HUMAN BIOLOGY
a record of research

SEPTEMBER, 1950

VOL. 22 No. 3

SAMPLING VARIATION OF THE FIFTY PER CENT END-POINT, DETERMINED BY THE REED-MUENCH (BEHRENS) METHOD * †

BY MARIO PIZZI, M. D., D. P. H.

I. INTRODUCTION

A NUMBER of statistical methods have been developed in recent years to interpret the results obtained in assaying drugs, insecticides and so forth. One of the most widely used short methods was developed independently by Dragstedt and Lang (5), Behrens (1), and Reed and Muench (13). This estimate is here examined for its accuracy and precision and a formula developed for its standard error. Because it has come to be called the Reed-Muench method in this country, it will be so referred to in this paper.

When groups of subjects are exposed to different concentrations of a stimulus which produces a reaction, such as a toxin causing death, the proportion of reactors plotted against the dose usually gives an *S*-shaped

* From the Department of Biostatistics (Paper No. 261), School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Md.

† Acknowledgment is made to Dr. E. K. Marshall, Jr. for making freely available the experiments done in his laboratory and to Dr. J. T. Litchfield, Jr., for kindly allowing me the use of his calculations of the LD 50 by probits on these experiments.

curve. The statistical analysis involves the determination of the function describing this dose-response curve.

Symmetrical functions such as the integrated normal curve or the logistic have been used empirically for this purpose. It has been found that a log scale of doses usually increases the symmetry of the material which originally might be somewhat skew, and in such cases this transformation is the first step in the analysis. The advantages of the change in scale were early realized by Krogh and Hemmingsen (9), Trevan (14), and Gaddum (7), among others. This change not only simplifies the analysis, but also facilitates the interpretation of results such as the comparison of a product of unknown potency with a standard, since this involves the concept of a dilution or concentration of the standard.

In the absence of some rationalization of the process leading to one or the other of these curves, the choice becomes a matter of convenience since, as shown by Winsor (18), the logistic and normal curves are practically indistinguishable within fairly wide limits. The problem then becomes one of curve fitting and if either of these functions is used, only two constants, giving the center and the spread, need to be determined in order to describe the material tested.

Historically, the statistical treatment of this type of data employed the " minimal effective dose " or the " maximal non-lethal dose " rather than a centering constant. This approach failed to recognize the fact that the curves are asymptotic and consequently the extreme doses have indefinitely large errors of estimation. Thus if we obtain an estimate of the " minimal lethal dose," i. e. the minimum dose which will kill all the animals, using a given number of animals, a larger sample would tend to contain a still more resistant member, thus increasing the value previously obtained. Because of its relative stability, Trevan (15) in 1927 sponsored the use of the median lethal dose (LD 50), which is defined as the dose which will make half of the population react to the stimulus tested.

The normal curve as a description of the material is conveniently fitted using the transformation which makes the integrated normal curve linear. Gaddum (7) in 1933 proposed to transform each percentage to its normal equivalent deviation, defined as the abscissa of a normal curve of zero mean and unit variance corresponding to the accumulated percentage up to that point. Bliss (3, 4) used the units of normal deviation increased by 5, to avoid negative values, and called these units

"probits." With the use of a log scale as the abscissa and a probit or normal equivalent deviation scale as the ordinate, the observed points plot approximately as a straight line; the maximum likelihood or least squares solution for the parameters is then obtained and the LD 50 and its standard error estimated therefrom.

The assumption that the original material plots as a logistic has been used mainly by Berkson (2) and by Wilson and Worcester (16, 17). This method uses the fact that the "logit" or natural logarithm of $(1 - q_i)/q_i$ (where q_i is the observed percentage of mortality at dose i) is linear with respect to the log of the dose. An estimate of the parameters is obtained from this line and the LD 50 and its standard error calculated.

The methods just outlined may not be practical for many of the assays required in the laboratory because the quality of the material studied, due to small numbers or other limitations of the experiment, hardly justifies an elaborate analysis. A number of useful approximate solutions have been developed to meet the needs of the biological worker for a simple and fast method yielding a good approximation for the LD 50 and its standard error.

One of these is the cumulative sum method of Reed-Muench. This assumes that the animals which react to a given dose would also have reacted to bigger ones, and those which did not respond at a certain level would not have reacted at still lower levels.

Algebraically this cumulative proportion, P, of reactors (deaths) for a particular dose x_a in a set of dosages between x_1 and x_k is given by

$$P_{x_a} = \frac{\sum_{1}^{a} d_i}{\sum_{1}^{a} d_i + \sum_{a}^{k} s_i}$$

where d_i and s_i are deaths and survivors respectively at dose i. The LD 50 is obtained by interpolation between the doses bracketing the 50 per cent position. The charming simplicity of this method has been responsible for its widespread use. It should be recognized that it is a method for getting only the median dose, since the cumulative proportions at other doses are not estimates of the true proportions at these doses. This summation process smooths the original material in such a way that the cumulative rates continually increase (or decrease) with increasing dosage thus making the estimation process for LD 50 easier in many cases than that of graphically fitting the probit line to the

observed proportions as proposed by Litchfield and Fertig (11). It has certain disadvantages however:

1) It is obvious that if the dosages are not symmetrically spaced about the true LD 50 and the numbers of animals used at each dose are not similarly symmetrically spaced, the estimate of LD 50 will be biased (Irwin (8));

2) No formula for the standard error of LD 50 has been derived which can be obtained from the observations at hand.

With regard to the first issue it is proposed to examine how serious the bias is if the conditions of the method (symmetry, equal spacing and number of animals) are not fully met. We are concerned with this issue chiefly as it affects the total error of estimate rather than because the bias itself is intrinsically important. This is studied through an analysis of a series of theoretical samples where the true LD 50 is known and also through a study of the results obtained with the Reed-Muench method and those given by the probit analysis in a series of actual laboratory tests.

With regard to the second issue, a practical formula is developed to fill this gap. In doing this, use is made of the formula given by Gaddum (7) for this method:

$$\text{S. E.}_{\text{LD}50} = \sqrt{\frac{0.66\, h\, \sigma}{n}}$$

where h is the interval between dosages, σ is the standard deviation of the individual responses of the subjects tested and n is the number of animals per dose. As it stands, this can be used only when σ is either known or can be assumed to be the same as in earlier experiences with the same experiment. A procedure for getting an estimate of σ from the observed material will be presented.

II. THE MATERIAL ON WHICH THE REED-MUENCH METHOD WAS STUDIED

In order to get an estimate of the bias that the Reed-Muench method might give under certain conditions, as well as to study the sampling error of the estimate of LD 50, two sets of experiments, one theoretical and one observed, were used.

1. *Theoretical experiments.* It was assumed that the true proportion dying, p, was given by the logistic:

$$p = \frac{1}{1 + e^{-x}} \qquad (1)$$

where x is the dosage.[1] For this curve, LD 50 = 0, and the standard deviation of the individuals, σ, is 1.81, or $\pi/\sqrt{3}$ (Winsor (18)). This curve is shown graphically in Fig. 1, with the abscissa given on both the x and the x/σ scales.

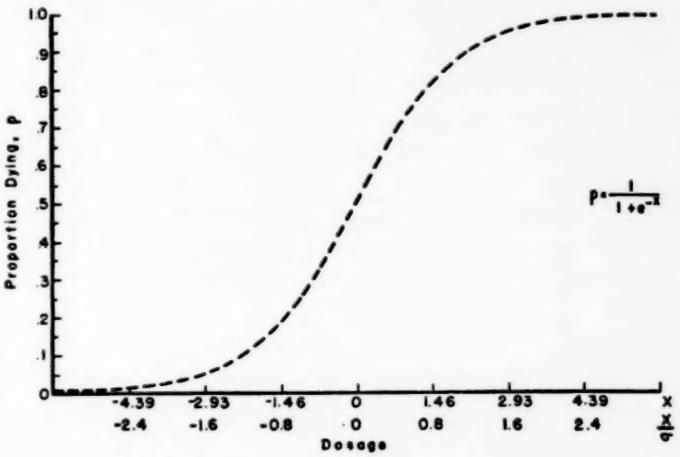

FIG. 1. THE LOGISTIC RESPONSE CURVE

A study was made of the sampling distribution of results when tests were made at four different doses equally spaced. The following factors in the situation were considered: (a) interval between dosages; (b) symmetry of dosages about LD 50; and (c) number of animals per dose.

Twenty-one theoretical experiments were conducted under different conditions as follows:

Symmetrical cases

Exp. no.	Interval h_x	Interval $h_{x/\sigma}$	Proportion dying at first dose	Number of animals per dose
1	0.7324	0.40	0.25	1
2	1.0986	0.60	0.16	1
3	1.1718	0.65	0.15	1
4	1.4648	0.80	0.10	1
9	1.7578	0.97	0.07	1
10	1.8310	1.00	0.06	1
4	1.4648	0.80	0.10	1
5	1.4648	0.80	0.10	2
6	1.4648	0.80	0.10	3
7	1.4648	0.80	0.10	4
8	1.4648	0.80	0.10	5

[1] Dosage will be used for that scale which gives a symmetrical response curve, usually the log of the original scale.

Asymmetrical cases

Exp. no.	Interval h_x	Interval $h_{x/\sigma}$	Proportion dying at first dose	Number of animals per dose
11	1.0986	0.60	0.10	1
12	1.0986	0.60	0.15	1
13	1.0986	0.60	0.19	1
14	1.4648	0.80	0.15	1
15	1.4648	0.80	0.15	2
16	1.4648	0.80	0.19	1
17	1.4648	0.80	0.19	2
18	1.4648	0.80	0.25	2
19	1.4648	0.80	0.32	2
20	1.4648	0.80	0.41	2
21	1.4648	0.80	0.59	2

In each experiment all the possible combinations of deaths and survivals were calculated. For instance, when there is one animal per dose, the possibilities are only that it can die or it can survive and with 4 doses the number of combinations is 2^4. In general, the number of combinations with n animals per dose and 4 doses is $(n+1)^4$. Thus, in the 21 experiments, the total number of different combinations of deaths and survivals is $11(2^4) + 7(3^4) + 1(4^4) + 1(5^4) + 1(6^4)$ or 2920. For each of these, except for certain indeterminate cases to be noted later, the LD 50 was obtained by the Reed-Muench method and the corresponding relative frequency was calculated by the suitable products of the independent probabilities given by the expansion of $(p+q)^n$ where p is the value given by equation (1) at each level, q is $1 - p$, and n is the number of animals per dose.[2]

---

[2] An example of the computation is given here for a particular sample in experiment 5, where: $p_1 = 0.10$; $p_2 = 0.32$; $p_3 = 0.68$; $p_4 = 0.90$ and $n = 2$. The different probabilities of occurrence of deaths and survivals at each dose are given by:

		2d	*1d, 1s*	*2s*
1st dose	$(p_1 + q_1)^2 =$	0.0100 +	0.1800 +	0.8100
2nd dose	$(p_2 + q_2)^2 =$	0.1024 +	0.4352 +	0.4624
3rd dose	$(p_3 + q_3)^2 =$	0.4624 +	0.4352 +	0.1024
4th dose	$(p_4 + p_4)^2 =$	0.8100 +	0.1800 +	0.0100.

With these frequency distributions,[3] the mean and the standard deviation of the LD 50's for each experiment were calculated by the usual procedure.

2. *Observed experiments.* The experimental data were obtained by testing the effect of seven different sulfanilamide derivatives in the presence of an infection which was invariably fatal to untreated mice (β-hemolytic streptococci mouse-virulent Group A). Details concerning the experiment are given in (10) and (12). Fifty-seven different tests were conducted over a two-year period, the results of which, as well as the number of animals and the dosage used, are given in the Appendix. The number of mice in any one experiment varied from 27 to 120. The number of animals used at any particular dose varied from 3 to 12; the number of levels varied from 4 to 12, most of the drugs being tested with 5 to 6 doses.

This material was analyzed by the probit method by Dr. J. T. Litchfield, Jr., and he discarded 3 experiments either because only 0% and 100% results were obtained or due to the great variability of the observations at certain levels. We discarded 4 more (exps. 6, 43, 52, 55) because the 50% value fell outside the range of the cumulative percentages as determined by the Reed-Muench method. Since this method has been recommended only for those cases where the LD 50 was either observed or bracketed, this study was limited to these instances.

III. THE LD 50 CALCULATED BY THE REED-MUENCH METHOD

1. *Theoretical experiments.* In calculating the LD 50 for the individual theoretical samples, each of the 21 experiments contained some samples for which the LD 50 was not determined because all animals

---

Then for the particular combination:

	d	*s*	*Probability*
1st dose	0	2	0.8100
2nd dose	1	1	0.4352
3rd dose	1	1	0.4352
4th dose	2	0	0.8100

the total probability of occurrence of this combination is equal to the product of the independent probabilities, i. e.

$$(0.8100)^2 \ (0.4352)^2 = 0.1243.$$

[3] The frequency distributions of these experiments are available in the files of the Department of Biostatistics of the School of Hygiene and Public Health, The Johns Hopkins University.

died in the sample, none died, or the 50% kill fell outside the range of the cumulative percentages. The situation of indeterminate values arises with any method used to calculate the LD 50 and others have called attention to this difficulty. It is apparent that some research should be done as to the best way of handling these undetermined values, since discarding them affects the estimate of the true standard error.

An illustration of the Reed-Muench procedure to determine the LD 50 is given in Table 1.

TABLE 1

Calculation of LD 50 by the Reed-Muench method

(Sample from experiment 15, true LD 50 = 0, $n = 2$.)

DOSAGE x	THEOR. p	DEAD d	LIVING s	ACCUMULATED DEATHS	ACCUMULATED SURV.	TOTAL	P (R. M.)
— 1.7346	0.15	1	1	1	4	5	0.20
— 0.2698	0.43	1	1	2	3	5	0.40
+ 1.1950	0.77	2	0	4	2	6	0.67
+ 2.6598	0.93	0	2	4	2	6	0.67

$$\text{LD 50 (Reed-Muench)} = -0.2698 + \frac{1.4648\,(0.10)}{0.27} = 0.27$$

Having obtained the LD 50 (x_{50}) for every sample and its probability of occurrence (f) as previously described, the mean x_{50} for each of the 21 experiments was calculated as $\Sigma f x_{50}/\Sigma f$. Since the sampling was done from a logistic with mean at 0, the bias of the method is given by the comparison of the calculated means with 0. The results are presented in Tables 2 and 3 where the first seven columns give the description of the experiment; column 8 shows the total proportion of cases where the LD 50 could be determined; columns 9 and 10 give respectively the sums necessary for computation of the mean and standard deviation of the distribution; column 11 gives the means of the distributions which are to be compared with 0, and column 12 enables us to judge how large the bias is by giving the true percentage dying corresponding to the mean of the distribution.

Considering first the results of Table 2, where the symmetrical cases are presented, it is seen from columns 8, 9, 11, and 12 that:

a) The proportion of indeterminate LD 50's decreases with increasing interval (exps. 1, 2, 3, 4, 9, 10). With an increasing number of animals tested at a dose, the indeterminate results als. lecrease (exps. 5, 6, 7, 8). For instance, from column 8 it is seen that the proportion of undeter-

TABLE 2

Summary of the distribution of median dosage in 10 theoretical sampling experiments from a logistic dosage-response curve—symmetrical cases

EXP. NO.	INTERVAL BETWEEN DOSAGES	NUMBER OF ANIMALS PER DOSE	TRUE PROPORTION AT LEVEL x_i				PROPORTION OF DETERMINATE x_{50}'s	MOMENTS OF DISTRIBUTION OF x_{50}		MEAN OF DIST. OF x_{50}	TRUE PROP. AT $m_{x_{50}}$
	h	n	p_1	p_2	p_3	p_4	Σf	$\Sigma f x_{50}$	$\Sigma f(x_{50})^2$	$m_{x_{50}}$	p_m
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	0.7324	1	0.25	0.41	0.59	0.75	0.9093	0	0.3407	0	0.50
2	1.0986	1	0.16	0.37	0.63	0.84	0.9373	0	0.6770	0	0.50
3	1.1718	1	0.15	0.36	0.64	0.85	0.9412	0	0.7549	0	0.50
4	1.4648	1	0.10	0.32	0.68	0.90	0.9608	0	1.0660	0	0.50
9	1.7578	1	0.07	0.29	0.71	0.93	0.9732	0	1.4141	0	0.50
10	1.8310	1	0.06	0.29	0.71	0.94	0.9768	0	1.5127	0	0.50
4	1.4648	1	0.10	0.32	0.68	0.90	0.9608	0	1.0660	0	0.50
5	1.4648	2	0.10	0.32	0.68	0.90	0.9927	0	0.7192	0	0.50
6	1.4648	3	0.10	0.32	0.68	0.90	0.9991	0	0.4831	0	0.50
7	1.4648	4	0.10	0.32	0.68	0.90	0.9995	0	0.3618	0	0.50
8	1.4648	5	0.10	0.32	0.68	0.90	0.9999	0	0.2874	0	0.50

UNIVERSITY OF MICHIGAN LIBRARIES

TABLE 3

Summary of the distribution of median dosage in 11 theoretical sampling experiments from a logistic dosage-response curve—asymmetrical cases

EXP. NO.	INTERVAL BETWEEN DOSAGES	NUMBER OF ANIMALS PER DOSE	TRUE PROPORTION AT LEVEL x_i				PROPORTION OF DETERMINATE x_{50}'s	MOMENTS OF DISTRIBUTION OF x_{50}		MEAN OF DIST. OF x_{50}	TRUE PROP. AT $m_{x_{50}}$
	h	n	p_1	p_2	p_3	p_4	Σf	$\Sigma f x_{50}$	$\Sigma f(x_{50})^2$	$m_{x_{50}}$	p_m
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
11	1.0986	1	0.10	0.25	0.50	0.75	0.9062	— 0.2094	0.6525	— 0.2311	0.44
12	1.0986	1	0.15	0.35	0.61	0.83	0.9368	— 0.0349	0.6763	— 0.0373	0.49
13	1.0986	1	0.19	0.41	0.68	0.86	0.9330	0.0647	0.6708	0.0694	0.52
14	1.4648	1	0.15	0.43	0.77	0.93	0.9460	0.1316	1.0295	0.1391	0.53
15	1.4648	2	0.15	0.43	0.77	0.93	0.9900	0.0601	0.6740	0.0607	0.52
16	1.4648	1	0.19	0.50	0.81	0.95	0.9230	0.2084	0.9700	0.2258	0.56
17	1.4648	2	0.19	0.50	0.81	0.95	0.9807	0.0933	0.6463	0.0952	0.52
18	1.4648	2	0.25	0.59	0.86	0.96	0.9584	0.1709	0.5987	0.1782	0.54
19	1.4648	2	0.32	0.68	0.90	0.97	0.9197	0.2555	0.5423	0.2779	0.57
20	1.4648	2	0.41	0.75	0.93	0.98	0.8541	0.3574	0.5167	0.4184	0.60
21	1.4648	2	0.59	0.86	0.96	0.99	0.6677	0.5487	0.6378	0.8218	0.69

mined LD 50's, when the interval between doses is 0.73 (exp. 1), is 0.09; doubling the interval (exp. 4), this proportion decreases to 0.04. This same experiment (4) with the number of animals per dose increased from 1 to 5 (exp. 8) shows a negligible loss of frequencies (0.0001) due to undetermined LD 50's.

b) It is clear that the values of LD 50 at equal distances either side of 0 have the same frequency, since they result from a simple reversal of p's and q's. Consequently, the frequency distributions are symmetrical about 0, the mean is 0, and the method thus has no bias when the doses are symmetrically spaced about the median with a fixed number of animals per dose (column 11).

If we consider now Table 3 where the asymmetrical experiments are presented, it is apparent that:

a) The indeterminate cases increase with the increase of asymmetry. In the most extreme case considered (exp. 21), where the sampling is all done at doses above the 50% value, the LD 50 could be obtained in only 67% of the cases (column 8).

b) On the other hand, the $\Sigma f(x_{50})$ increases with asymmetry (column 9), and this fact combined with the decreasing frequency of values, increases the mean, resulting in the bias shown in columns 11 and 12. For a given interval and degree of asymmetry, the bias decreases as the number of animals per level increases (exps. 14 and 15; 16 and 17). Thus in experiment 16, with one animal per level, the mean LD 50 corresponds to a true death rate of 56%, while in experiment 17, which differs only by having two animals per dose, the mean corresponds to a death rate of 52% (column 12). If the asymmetry had been taken in the other direction from the mean, obviously the corresponding experiments would have the same bias on the negative side of the curve.

The effect on the bias of those combinations where the LD 50 is not bracketed but falls exactly at one of the extreme levels (the first or the last dose tested) was also considered. If these cases are eliminated from the calculations it is found that the bias is further increased. In an actual experiment, therefore, the Reed-Muench method may be applied not only when the LD 50 is bracketed, but also when it falls at one of the extreme doses, without fear of increasing the bias by this inclusion. This raises the question as to whether the discarding, as has been done in this study, of those cases where the LD 50 could only have been obtained by extrapolation is desirable since it may increase the bias enough to increase the error of estimate.

In practice it is possible to plan an experiment with equally spaced doses and an equal number of animals at each dose, but symmetry about the unknown mean cannot be assured. The tests just analyzed indicate that the bias of the method is negligible when the two central doses straddle the 50% point even with as small numbers as 1 or 2 at a dose and that it decreases with increasing size of sample. This agrees with the conclusion of Irwin (8). For the case of quite extreme skewness where all doses tested were above the median, and where the central position of the doses tested corresponded to a true mortality of 90% the average value of those LD 50's which could be determined corresponded to a mortality of 69% when there were as few as two animals to a dose. The seriousness of the bias as it affects the standard error of estimate will be considered in the next section.

2. *Observed experiments.* For each of the 50 observed laboratory experiments previously described, the LD 50 obtained by the Reed-Muench method was compared with that calculated by probit analysis. The results can be seen in Table 4 and Fig. 2 which relates the two estimates.

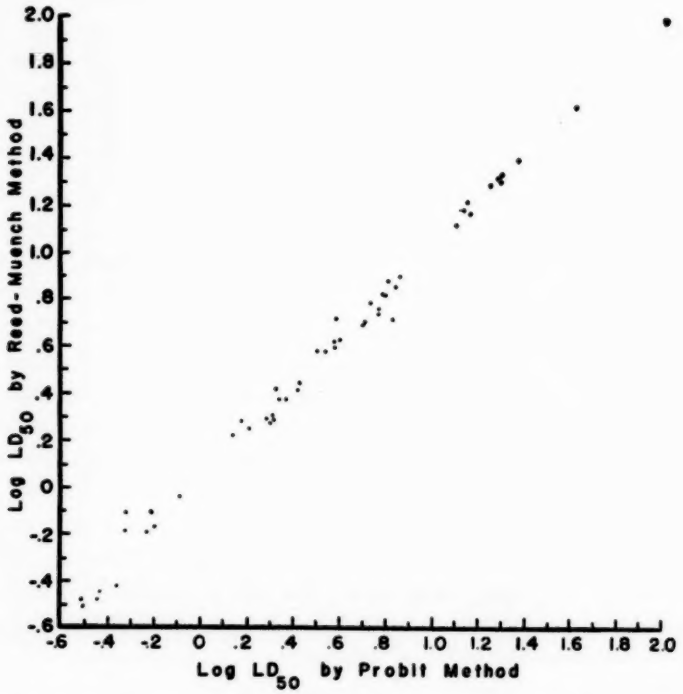

FIG. 2. CORRELATION BETWEEN LOG LD 50 OBTAINED BY PROBIT AND REED-MUENCH METHODS ON 50 OBSERVED EXPERIMENTS

TABLE 4

The LD 50 calculated by Reed-Muench and probit methods in 50 observed experiments

(Values on log scale.)

EXP. NO.	LD 50 PROBIT x	LD50 REED-MUENCH y	$\frac{x-y}{\text{S. E.}^{*}\text{LD50}}$	EXP. NO.	LD 50 PROBIT x	LD50 REED-MUENCH y	$\frac{x-y}{\text{S. E.}^{*}\text{LD50}}$
(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
1	0.428	0.441	0.29	28	0.787	0.820	0.79
2	0.318	0.292	0.30	29	0.709	0.698	0.14
3	0.426	0.415	0.12	30	— 0.209	— 0.109	1.50
4	0.341	0.377	0.66	31	1.624	1.615	0.16
5	0.143	0.223	0.77	32	1.159	1.210	1.16
7	0.293	0.288	0.08	34	1.291	1.311	0.42
8	0.318	0.305	0.28	35	1.300	1.304	0.01
9	— 0.505	— 0.511	0.06	36	0.211	0.255	1.38
10	— 0.439	— 0.479	0.39	37	1.374	1.391	0.40
11	— 0.507	— 0.479	0.33	38	1.253	1.284	0.50
12	— 0.359	— 0.423	0.43	39	1.300	1.333	0.42
13	0.372	0.378	0.02	40	0.177	0.286	1.16
14	0.302	0.279	0.50	41	2.005	1.968	0.92
15	— 0.433	— 0.453	0.38	42	1.162	1.163	0.02
16	0.801	0.815	0.14	44	0.511	0.577	0.74
17	0.736	0.779	0.61	45	0.603	0.626	0.29
18	0.864	0.893	0.55	46	1.102	1.111	0.09
19	0.810	0.872	0.70	47	0.704	0.687	0.21
20	0.832	0.707	0.54	49	0.580	0.611	0.32
21	0.769	0.729	0.24	50	1.137	1.176	0.63
22	0.770	0.748	0.18	51	0.546	0.577	0.36
23	0.584	0.594	0.10	53	0.594	0.708	0.37
24	— 0.325	— 0.187	0.67	54	0.840	0.852	0.04
25	— 0.227	— 0.195	0.24	56	— 0.198	— 0.175	0.21
27	— 0.093	— 0.043	0.59	57	— 0.320	— 0.119	0.72

* Standard error from probit analysis.

The mean difference between the results (probit — Reed-Muench) for these experiments is -0.0213 ± 0.007 which corresponds to a difference/S. E.$_{diff.}$ of 3. This result studied in the light of the range of values in a rather homogeneous set of experiments, indicates that the

mean difference between the LD 50's obtained by the two methods, though significant, is of small order. The ratio

$$\frac{\text{LD 50 (probit)} - \text{LD 50 (Reed-Muench)}}{\text{S. E.}_{\text{probit LD 50}}}$$

is given in column 4, Table 4 and is summarized in Table 5. It is seen

TABLE 5

Differences between LD 50's obtained by Reed-Muench and probit methods expressed in S. E. probit LD 50 units

LD 50 (PROBIT) — LD 50 (R. M.) / S. E. probit LD 50	FREQUENCY
+ 0.00-0.49	31
+ 0.50-0.99	15
+ 1.00-1.49	3
+ 1.50-1.99	1
Total	50

that 92% of the LD 50's as obtained by the Reed-Muench method are less than one standard error (probit) from the corresponding LD 50's given by probits, and that the remaining are between one and two standard errors. The mean of this distribution is 0.5.

This agreement is of special importance to the investigator, since these experiments do not fully meet the theoretical conditions for the application of the Reed-Muench method because in most of them the interval between doses is not exactly the same and in some of them the number of animals per level is not constant. Which of the two methods makes more complete use of the information of the experiment is not at present known, because the efficiency of the probit method for small samples has not been determined.

IV. THE STANDARD ERROR OF LD 50 IN THE THEORETICAL EXPERIMENTS

1. *General effects of sample size, width of spacing and asymmetry.* Before presenting an analytical approach to the standard error of the Reed-Muench LD 50, the variation in the 21 theoretical distributions is examined in order to assess the importance of the bias in the error of estimate. Only the standard deviation about the true mean is presented since we are concerned with the deviations of our estimates

from this value. This standard deviation is compared with that given by Gaddum for the Behrens (Reed-Muench) LD 50, so that the validity of this formula may be studied. In addition, the standard error for the conditions of these experiments is computed from a formula furnished by Wilson and Worcester (16, 17) for the standard error from the maximum likelihood solution for the logistic, this being a method virtually equivalent to that of probits.

TABLE 6

The standard deviations of the LD 50 about the true mean compared with the standard error obtained by Gaddum and Wilson-Worcester formulas in 10 theoretical experiments—symmetrical cases

EXP. NO.	INTERVAL BETWEEN DOSES h	NUMBER OF ANIMALS PER DOSE n	TRUE PROP. AT FIRST LEVEL p_1	ST. DEV. ABOUT TRUE MEAN (0) σ LD50	S. E. OF LD50	
					Gaddum	W. W.
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0.7324	1	0.25	0.612	0.936	1.079
2	1.0986	1	0.16	0.850	1.147	1.166
3	1.1718	1	0.15	0.896	1.184	1.182
4	1.4648	1	0.10	1.053	1.324	1.275
9	1.7578	1	0.07	1.205	1.451	1.358
10	1.8310	1	0.06	1.244	1.481	1.381
4	1.4648	1	0.10	1.053	1.324	1.275
5	1.4648	2	0.10	0.851	0.936	0.902
6	1.4648	3	0.10	0.695	0.765	0.736
7	1.4648	4	0.10	0.602	0.662	0.637
8	1.4648	5	0.10	0.536	0.592	0.570

The results for the symmetrical cases are presented in Table 6 and for the asymmetrical in Table 7, where

Columns 2-4 give the description of the experiment,

Column 5 gives the standard deviation of LD 50 about the true mean,

Column 6 gives the standard error of LD 50 as obtained by Gaddum's formula, and

Column 7 gives the standard error of LD 50 as obtained by the Wilson-Worcester formula.

TABLE 7

The standard deviations of the LD 50 about the true mean compared with the standard error obtained by Gaddum and Wilson-Worcester formulas in 11 theoretical experiments—asymmetrical cases

NO. EXP.	INTERVAL DOSES BETWEEN h	NUMBER OF PER DOSE ANIMALS n	TRUE PROP. LEVEL AT FIRST p_1	TRUE MEAN (0) ST. DEV. ABOUT σ_{LD50}	S. E. OF LD 50	
					Gaddum	W. W.
(1)	(2)	(3)	(4)	(5)	(6)	(7)
11	1.0986	1	0.10	0.849	1.147	1.183
12	1.0986	1	0.15	0.850	1.147	1.167
13	1.0986	1	0.19	0.848	1.147	1.167
14	1.4648	1	0.15	1.043	1.324	1.275
15	1.4648	2	0.15	0.825	0.936	0.902
16	1.4648	1	0.19	1.025	1.324	1.285
17	1.4648	2	0.19	0.812	0.936	0.909
18	1.4648	2	0.25	0.790	0.936	0.922
19	1.4648	2	0.32	0.767	0.936	0.950
20	1.4648	2	0.41	0.778	0.936	0.986
21	1.4648	2	0.59	0.977	0.936	1.104

For the symmetrical cases, the standard deviation about the true mean (which in this case is also the mean of the distribution) increases with the interval between doses (column 6) when experiments with the same number of animals per dose are considered (exps. 1, 2, 3, 4, 9, 10). For a particular interval, the standard deviation decreases as n increases (exps. 5, 6, 7, 8) in the way predicted by theory, i. e. according to $1/\sqrt{n}$.

The behavior of the standard deviations for the asymmetrical cases presented in Table 7, is somewhat more involved. On general grounds, it might be expected that increasing the asymmetry would have the effect of increasing the standard deviation, due to the extension of one of the tails. This has to be balanced against the effects of the concentration on one side, as may be seen in Fig. 3, where the relative frequencies of LD 50's in experiments 5, 17, 19 and 20 are presented. The standard deviation shows a minimum value which is not at the symmetrical position. In Fig. 4, the standard deviations corresponding to experiments 5, 15, 17, 18, 19, 20 and 21 have been plotted, together with those pertaining to corresponding asymmetrical experiments in the other direction. The means of the distributions of LD 50, plotted on

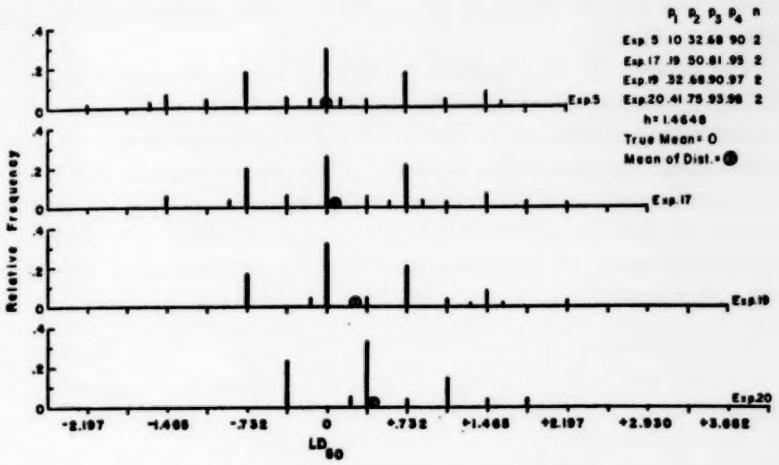

FIG. 3. RELATIVE FREQUENCY DISTRIBUTION OF LD 50 IN SPECIFIED SYMMETRICAL AND ASYMMETRICAL THEORETICAL EXPERIMENTS

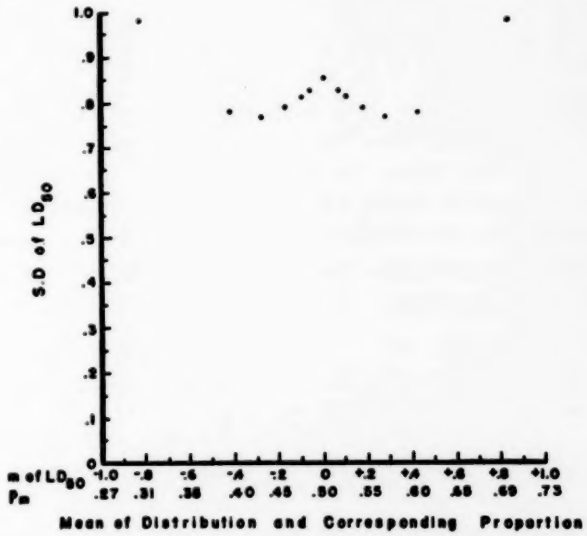

FIG. 4. STANDARD DEVIATION OF LD 50 ABOUT THE TRUE MEAN IN THEORETICAL EXPERIMENTS WHERE $h = 1.46$ AND $n = 2$ AGAINST THE MEAN OF THE DISTRIBUTION (EXPERIMENTS 5, 15, 17, 18, 19, 20, AND 21)

the abscissa are an index of asymmetry, with 0 representing the symmetrical position.

The pattern of the standard deviations can be explained in part by the loss of frequencies due to the undetermined LD 50's and in part by the increasing skewness of the distributions. Tables 2 and 3 for the experiments just cited show that the $\Sigma f(x_{50})^2$ decreases steadily up to experiment 20 and this decrease is proportionally bigger than the decrease of Σf. In experiment 20, on the other hand, the decrease of Σf is proportionally greater than that of $\Sigma f(x_{50})^2$ and thus the standard deviation increases.

2. *Contribution of bias to the standard error.* The effects of the bias on the standard error can be studied through a comparison of the standard deviation about the true mean and that about the mean of the distribution, which is not shown here but may be obtained from columns 9 and 10 of Table 3. Analysis shows that the influence of the bias on the standard deviation is negligible (less than one per cent contribution to the standard error) when the LD 50 is straddled by the two central doses (exps. 13, 14, 15). When the 50% position is hit by the second or the third dose, the contribution of the bias to the error of estimate varies from 2 to 4% (exps. 11, 16, 17). If the 50% position falls between the first and second dose, (or the third and fourth) the contribution depends on how close the 50% value is to the extreme dose. Thus, where the first dose carries a mortality of 32%, the bias contributes 7% to the standard error, and where the lowest dose was as high as 41% the contribution is 16%. When all 4 of the doses are above the 50% value, the lowest being at 59%, the bias contributes 46% to the error of estimate.

3. *Comparison with the Gaddum and Wilson-Worcester formulas.* The standard deviation of LD 50 for these experiments was compared with that estimated by Gaddum's formula:

$$\text{S. E.}_{\text{LD}\,50} = \sqrt{\frac{0.66\, h\, \sigma}{n}} \qquad (2)$$

which was empirically obtained for the present method by an actual sampling from the normal curve with standard deviation σ. In the experiments considered, this formula becomes

$$\text{S. E.}_{\text{LD}\,50} = \sqrt{\frac{1.20\, h}{n}},$$

since (Winsor, (18))

$$\sigma = \pi/(b\sqrt{3}), \tag{3}$$

and b, the coefficient of x in the logistic, in this case is 1 from equation (1). The standard errors obtained by this formula are presented in column 6 of Tables 6 and 7. It is seen that they are constantly higher than those calculated directly from the distributions, except for experiment 21. It was pointed out that the directly determined standard deviations were lowered somewhat by the loss of frequencies corresponding to undetermined LD 50's which would represent the more extreme cases, and this would explain some of the difference between the two values. However, even in the case where the frequency of undetermined LD 50's is negligible, Gaddum's formula gives results about 10% higher than those determined from the distributions. Considered against the variation in the standard error of LD 50 obtained in actual practice resulting from estimating σ from the observed material, the difference seems unimportant and has the merit of giving an estimate of the confidence limits which is on the safe side. It is possible that the numerical coefficient in Gaddum's equation is slightly too high, but we have considered that the agreement was sufficiently satisfactory to allow the use of this formula as an estimate of the standard error.

It is of interest to compare these values with those given by the Wilson-Worcester formula:

$$\text{S. E.}_{\text{LD}\,50} = \frac{1}{b\sqrt{n\Sigma p_i q_i}}. \tag{4}$$

This is the maximum likelihood solution for the standard error of LD 50 assuming that the original material is a logistic response curve of known b. As indicated in equation (3), b is a measure of the variability of the individual responses of the subjects tested (unity in our case); p and q are the true proportions dying and surviving, as given by the logistic, and n is the number of animals per dose. The results are presented in column 7 of Tables 6 and 7. It is seen that they are very similar to those given by Gaddum's formula, in general differing by about 4% and at most by about 15%. It is interesting that they agree so well since they pertain to estimates of LD 50 obtained by quite different procedures. It should be noted, however, that the standard error as obtained by the Wilson-Worcester formula increases with asymmetry while the one given by Gaddum does not provide for such an adjustment.

Obviously, in this instance we were able to apply these formulas because we knew b and the true proportions dying. This is not the situation in an actual experiment and it is necessary to estimate b from the observed material.

V. ESTIMATE OF THE STANDARD DEVIATION FROM THE OBSERVATIONS

1. *An approximation to the cumulative Reed-Muench curve.* Since the standard deviation of the individuals is, from equation (3), a function of the rate constant, b, in the logistic assumed to describe the dose response curve, i. e.

$$f(x) = \frac{1}{1 + e^{a-bx}}, \tag{5}$$

the derivation of b leads to the standard error of LD 50. It was therefore desired to get the relationship between the increase for the Reed-Muench cumulative percentages and the b constant. The difficulty of dealing with the finite summation process analytically, led to a consideration of the function defined in an analogous way by integration. The function thus defined is:

$$F(x) = \frac{\int_{-\infty}^{x} f(x)\,dx}{\int_{-\infty}^{x} f(x)\,dx + \int_{x}^{\infty} [1 - f(x)]\,dx}. \tag{6}$$

Solving, we get

$$F(x) = \frac{a - bx - \log_e(1 + e^{a-bx})}{a - bx - 2\log_e(1 + e^{a-bx})}. \tag{7}$$

It is easily shown that when $x \to -\infty$, $f(x)$ and $F(x)$ approach 0;

when $x = a/b$, $f(x)$ and $F(x) = 0.5$;

and when $x \to +\infty$, $f(x)$ and $F(x)$ approach 1;

and also that both curves are symmetrical about $x = a/b$, i. e. x_{50}.

The curves $F(x)$ and $f(x)$ are plotted in Fig. 5 for the case when $a = 0$ and $b = 1$. It is seen that $F(x)$ is similar to $f(x)$ in form, but steeper at the 50% position. In Fig. 6,

$$\frac{1 - f(x)}{f(x)} \quad \text{and} \quad \frac{1 - F(x)}{F(x)}$$

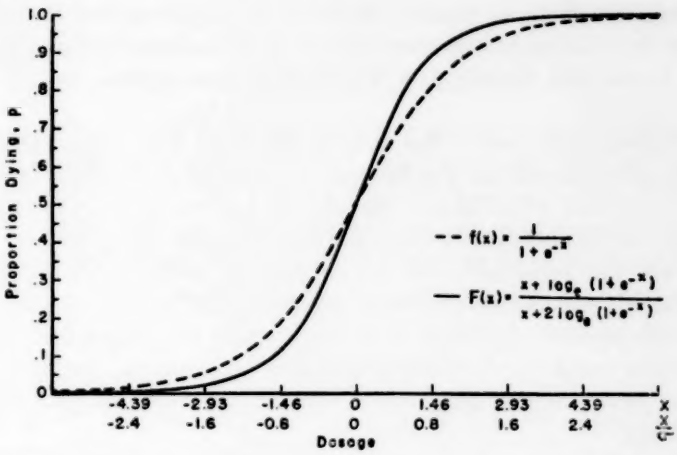

FIG. 5. THE LOGISTIC RESPONSE CURVE, $f(x)$, AND THE INTEGRAL FUNCTION OF THE LOGISTIC, $F(x)$

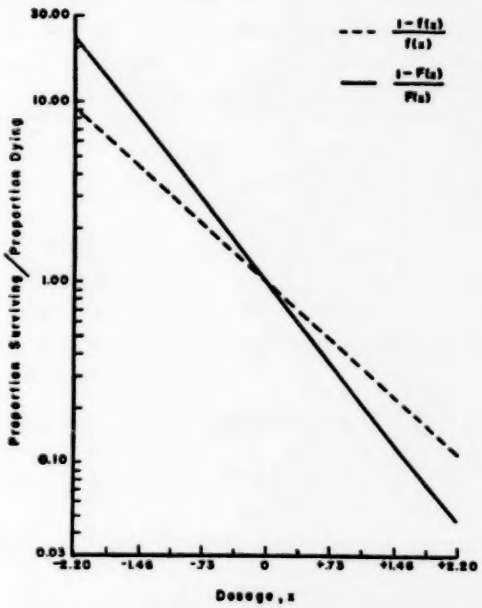

FIG. 6. LOG OF PROPORTION SURVIVING/PROPORTION DYING FROM $f(x)$ AND $F(x)$ AGAINST DOSAGE

have been plotted on log paper. It will be remembered that the logistic is linear for this transformation and it is seen that $F(x)$ is approximately linear and therefore is approximately a logistic, over a wide range.

To indicate how closely the integral function, $F(x)$, approaches the Reed-Muench summation proportions, the log of $(1 - P)/P$ for each of the theoretical experiments was plotted against the dosage together with the corresponding functions of the $f(x)$ and $F(x)$ curves. The results are not presented here but it may be easily verified that the Reed-Muench summation process is approximated by the corresponding integration process. Different intervals, number of dosages, and degrees of skewness result in different cumulative percentages, but the curve resulting from integration forms a sort of core for the Reed-Muench summations.

2. *Relation between the slopes of $F(x)$ and $f(x)$.* The slopes of $f(x)$ and $F(x)$ at the 50% position are functions of b, and their relationship permits the estimation of the b constant of the basic logistic from the cumulative curve. If the latter is considered to be approximated by the Reed-Muench percentages, we can through these and the slope relationship get an estimate of the standard error.

We need then, the first derivatives of $f(x)$ and $F(x)$ with respect to x, as follows:

$$\frac{df(x)}{dx} = \frac{be^{a-bx}}{(1 + e^{a-bx})^2}, \text{ and} \tag{8}$$

$$\frac{dF(x)}{dx} = \frac{b\left[\log_e(1 + e^{a-bx}) - \dfrac{(a - bx)e^{a-bx}}{1 + e^{a-bx}}\right]}{[a - bx - 2\log_e(1 + e^{a-bx})]^2}. \tag{9}$$

When $x = a/b$, i. e. at the 50% point:

$$\frac{df(x)}{dx} = \frac{b}{4} = 0.25b,$$

$$\frac{dF(x)}{dx} = \frac{b \log_e 2}{(-2 \log_e 2)^2} = 0.36b,$$

and the ratio between the two slopes at the 50% position is then

$$0.36b/0.25b = 1.44. \tag{10}$$

Consequently, the slope of the integral function $F(x)$, when this function has the value 0.5, is 1.44 times the slope of the original logistic

at the same point. Thus the slope at the 50% point of the percentages obtained by the summation process is also approximately 1.44 times the slope of the logistic response curve that originated them.

3. *An estimate of the slope of $F(x)$ from the data in an experiment.* To get an estimate of the slope of $F(x)$ from the observed material, we may use the fact that $F(x)$ is approximately a logistic over a substantial range. We may therefore write as an approximation to $F(x)$, the logistic which has the same coordinates and slope at the 50% value as $F(x)$. Designating this as P_x, we have

$$P_x = \frac{1}{1 + e^{1.44(a-bx)}}. \tag{11}$$

To get an estimate of b we may take the solution given by any two points r units apart, which is:

$$b = \frac{0.69}{r} \log_e \frac{Q_x P_{x+r}}{P_x Q_{x+r}} \tag{12}$$

$$= \frac{1.59}{r} \log_{10} \frac{Q_x P_{x+r}}{P_x Q_{x+r}}, \tag{13}$$

where $Q_x = 1 - P_x$. Using as estimates of P_x and P_{x+r}, any two suitable values of the observed Reed-Muench proportions, we have a method of estimating b. No definite rule can be given as to the choice of these points, but it is desirable to avoid selecting a point which, due to sampling variation, is out of the general trend of the experiment.

A convenient pair of points are the quartiles, in which case r represents the interquartile range, R. The dosages corresponding to the 0.25 and 0.75 positions may be obtained by interpolation, or if both points are not bracketed, the interquartile range can be estimated as twice $x_{75} - x_{50}$ or $x_{50} - x_{25}$.

4. *The formula for the standard error of LD 50.* If we put $P_x = 0.25$ and $P_{x+r} = 0.75$, then r becomes R and equation (13) can be written

$$b = \frac{1.59}{R} \log_{10} 9 = \frac{1.52}{R}. \tag{14}$$

Replacing this value in formula (3) for the standard deviation of the individual responses, we get

$$\sigma = 1.81\, R/1.52 = 1.19\, R. \tag{15}$$

Substituting this value for σ in Gaddum's formula (2), we have

$$\text{S. E.}_{\text{LD 50}} = \sqrt{\frac{0.79\, h\, R}{n}}\,. \qquad (16)$$

A numerical example of the application of this formula is given in Table 8 using the same material presented in Table 1.

TABLE 8

Calculation of the standard error of LD 50 using proposed formula
(Sample from experiment 15, true LD 50 = 0, $n = 2$.)

DOSAGE	DEAD	LIVING	P (R. M.)
x	d	s	
(1)	(2)	(3)	(4)
— 1.7346	1	1	0.20
— 0.2698	1	1	0.40
+ 1.1950	2	0	0.67
+ 2.6598	0	2	0.67

LD 50 was found to be 0.27
By interpolation we get $x_{25} = -1.37$
therefore $x_{50} - x_{25} = 1.64$

Since x_{75} can not be obtained in this case except by extrapolation, we estimate the interquartile range, R, as $2(1.64) = 3.28$; the interval between dosages (h) is 1.46 and n is 2. Substituting these values in (16) we get

$$\text{S. E.}_{\text{LD 50}} = \sqrt{\frac{0.79\,(1.46)\,3.28}{2}} = 1.38.$$

In summary, the reasoning in getting an estimate of the standard error of LD 50 has been as follows:

a) The original material is assumed to be described by a logistic response curve, $f(x)$;

b) The standard error of LD 50 is a function of the standard deviation of the individual responses, which is inversely proportional to the slope of $f(x)$ at the 50% point;

c) The Reed-Muench summation process is approximated by the analogous integral function $F(x)$, which is itself approximately a logistic with a slope at the 50% value 1.44 times that of $f(x)$;

d) The observed Reed-Muench percentages yield an estimate of the slope of $F(x)$ at the 50% position and thus of the slope of $f(x)$, and

the standard deviation of the individual responses. With this the standard error of LD 50 can be obtained from Gaddum's formula.

VI. PROPOSED FORMULA APPLIED TO THE OBSERVED EXPERIMENTS

Applying formula (16) to the 50 observed experiments referred to previously, we may compare the standard error of LD 50 with that given by the probit method. The results are presented in Table 9.

TABLE 9

The standard error of log LD 50 calculated by the probit and the proposed Reed-Muench formulas in 50 observed experiments

EXP. NO.	S. E. OF LOG LD50		EXP. NO.	S. E. OF LOG LD50	
	Probit x	Reed-Muench y		Probit x	Reed-Muench y
1	0.045	0.057	28	0.042	0.043
2	0.086	0.084	29	0.064	0.051
3	0.093	0.074	30	0.066	0.064
4	0.054	0.072	31	0.057	0.043
5	0.104	0.088	32	0.044	0.062
7	0.066	0.075	34	0.048	0.042
8	0.047	0.068	35	0.041	0.051
9	0.097	0.088	36	0.032	0.047
10	0.102	0.097	37	0.043	0.053
11	0.085	0.114	38	0.062	0.056
12	0.150	0.099	39	0.078	0.063
13	0.038	0.043	40	0.094	0.064
14	0.046	0.039	41	0.040	0.060
15	0.052	0.047	42	0.047	0.050
16	0.098	0.111	44	0.089	0.068
17	0.071	0.094	45	0.080	0.063
18	0.053	0.102	46	0.099	0.068
19	0.089	0.116	47	0.081	0.066
*20	0.233	0.114	49	0.096	0.058
21	0.165	0.157	50	0.062	0.060
22	0.124	0.110	51	0.086	0.078
23	0.099	0.099	*53	0.311	0.086
*24	0.205	0.137	54	0.133	0.083
25	0.134	0.145	56	0.108	0.083
27	0.085	0.080	*57	0.281	0.090

For most of the experiments the results are in good agreement. Four experiments, marked with an asterisk, differ markedly. Reference

to these in the Appendix shows them to be difficult experiments to interpret because of the irregularity of the observed mortalities. They are the kinds of experiments which are usually repeated. In two of them, the results given by probits seem questionable, and in the other two the Reed-Muench standard error seems too small. These are cases where the estimation of the standard deviation by any method gives highly uncertain results.

An analysis of variance between methods to test the significance of the difference of results gives the following values:

	D. F.	*Σ Sq.*	*Mean Sq.*	*F*
Total	**99**	**0.2136**		
Items	**49**	**0.1513**	**0.0031**	**2.66**
Methods	**1**	**0.0055**	**0.0055**	**4.76**
Discrepance	**49**	**0.0568**	**0.0012**	**1.00**

The F ratio for methods, 4.76 for 1 and 49 degrees of freedom, falls between the 1% (7.20) and the 5% (4.04) levels, close to the latter. The F ratio for discrepancies between methods, eliminating the 4 experiments previously cited, becomes 1.045, well above the 20% level of significance.

It is concluded that Gaddum's formula, with an estimate of σ through the interquartile range, as previously developed, gives an approximation of the standard error of LD 50 good enough for practical purposes.

If we want to get the standard error of LD 50 from the standard error of log LD 50, we may use the relationship that the standard error of a function of a variable is approximately equal to the first derivative of the function with respect to the variable times the standard error of the variable.

Thus if $$m = \log \text{LD } 50,$$

and $$u = \text{LD } 50,$$

then $$\text{S. E.}_{\text{LD}50} = (du/dm)\,\text{S. E.}_m,$$

or $$\text{S. E.}_{\text{LD}50} = 2.3(\text{LD } 50)\,\text{S. E.}_{\log \text{LD}50}.$$

It should be remembered that if the confidence limits of LD 50 are taken on the log scale, the upper and lower limits will be symmetrical with respect to log LD 50, but if they are taken on the original scale, they will no longer be symmetrical, and the upper and lower limits may be obtained by the antilog of the limits on the log scale.

VII. SUMMARY

The aims of this paper have been, first, to examine the Reed-Muench method for obtaining LD 50 as to the extent of its bias and the contribution of the bias to the total error of estimate, under various conditions of experimentation, and second, to develop a practical formula for an approximate solution of the standard error of LD 50.

An analysis of frequency distributions of the LD 50's obtained from theoretical sampling distributions under specified conditions indicated:

1) The method has no bias when the doses are symmetrically spaced about the LD 50 with equal intervals between dosages and a constant number of animals per level. This observation is not new and in fact is obvious from a consideration of the Reed-Muench process.

2) Sampling at 4 levels, for the intervals tested, the effect of the bias is negligible (less than 1% contribution to the standard error) when the LD 50 is straddled by the two central doses. This is true even though there are only one or two animals per dose. With increasing asymmetry, the contribution of the bias becomes more serious, and for the case where all 4 of the doses were above the 50% value (first dose at 59%, fourth at 99%) and there were 2 animals per dose, the bias contributed 46% to the standard error. Increasing the number of animals has the effect of decreasing the influence of the bias on the standard error of estimate.

A further examination of the Reed-Muench estimate of LD 50 is given by the comparison of the results of this method with those obtained by probit analysis on 50 observed experiments which involved from 27 to 120 animals per experiment. The comparison showed that none of the Reed-Muench LD 50's differed significantly from those of the probit method and that on the average they were away by 0.5 of the standard error given by the probit method. From our present knowledge for small samples there is no way to know which is more efficient, but the Reed-Muench process has the advantage of its extreme simplicity.

Gaddum's formula for the standard error of LD 50 for this method was applied to the theoretical experiments, and the results were compared with the standard deviations obtained directly from the distributions. It was found that Gaddum's formula gives slightly higher results than those determined from the distributions, due in part to the lowering of the latter by the loss of cases where LD 50 was undetermined. The

difference, however, seems unimportant when considered against the variation in the standard error of LD 50 obtained in actual practice resulting from estimating σ from the observed material.

In order to use Gaddum's formula in practice it is necessary to obtain an estimate of the standard deviation of the individual responses. For this purpose, it was assumed that the original material is described by a logistic response curve, $f(x)$. The integral function of the logistic, $F(x)$, analogous to the finite summation of the Reed-Muench process, was derived and was found to be a good description of the summation process. An estimate of the standard deviation of the individual responses was obtained from the interquartile range of the cumulative percentages, making use of the relationship between $F(x)$ and $f(x)$. This value, substituted in Gaddum's formula for the standard error of LD 50 gives our proposed formula:

$$\text{S. E.}_{\text{LD}\,50} = \sqrt{\frac{0.79\,h\,R}{n}}$$

where h is the interval between dosages, R is the interquartile range from the cumulative percentages, and n is the number of animals per dose. If the intervals are not equal and n is not constant, the corresponding averages are suggested as an estimate of h and n.

This formula was applied to 50 observed experiments and the results compared with those given by the probit analysis. The agreement was found good enough for practical purposes. A more precise interpretation of the differences observed calls for more research, since the efficiency of the probit method in small samples is not known.

On the basis of the analysis of the Reed-Muench method presented, it is recommended that in the planning of an experiment in bioassay the investigator should try to keep the interval between doses and the number of animals per level constant as far as possible. Symmetry should be sought even though it is recognized that it is difficult to achieve it.

Finally, attention should be called to the fact that the summation process involved in the Reed-Muench method has led some investigators to believe that it gives an accuracy to the percentage mortalities equivalent to that which would be obtained by the use of the actual number of animals that the summation gives for each level. This is not true. The cumulative percentages have only the effect of smoothing the observed results. As Behrens (1) pointed out: "The deduced ratio at

a given dose is not the mortality ratio except at the 0.5 value, since the ratios deal with areas on, rather than with ordinates of, the cumulative frequency curve." This means that the comparison of the relative potency of two drugs, for instance, through the slopes of the lines obtained by the cumulative percentages is not sound. The Reed-Muench method, as originally developed, can be used only to determine the LD 50 and not to calculate other lethal doses. However, through the approximation we have called $F(x)$, it is now possible to get estimates of doses corresponding to probabilties of dying other than the 50% value.

In the analysis of the theoretical experiments the bias of the Reed-Muench method in the asymmetrical distributions was studied. The direction of the bias raises an additional question to be investigated as to whether the limitations imposed in using the method, namely that the experiment should be discarded if the LD 50 is not bracketed or observed at one of the extreme doses, is a necessary and desirable limitation.

REFERENCES

1. Behrens, B. Zur Auswertung der Digitalisblätter im Froschversuch. *Arch. exp. Path. Pharmak.*, 140: 237-256, 1929.
2. Berkson, J. Application of the logistic function to bioassay. *J. Am. Stat. Ass.*, 39: 357-365, 1944.
3. Bliss, C. I. The method of probits. *Science*, 79: 38-39, 1934.
4. Bliss, C. I. The method of probits—a correction. *Science*, 79: 409-410, 1934.
5. Dragstedt, C. A. and Lang, V. F. Respiratory stimulants in acute cocaine poisoning in rabbits. *J. Pharmacol. and Exper. Therap.*, 32: 215-222, 1928.
6. Fisher, R. A. On the mathematical foundations of theoretical statistics. *Philos. Trans.*, 222: 309-368, 1922.
7. Gaddum, J. H. Reports on biological standards. III—Methods of biological assay depending on a quantal response. *Spec. Rep. Ser. Med. Res. Coun., London*, 183: 1-46, 1933.
8. Irwin, J. O. Statistical method applied to biological assay. *J. Roy. Stat. Soc. Suppl.*, 4: 1-60, 1937.
9. Krogh, A. and Hemmingsen, A. M. The assay of insulin on rabbits and mice. *Biol. Medd., Kbh.*, 7: 6, 1928.
10. Litchfield, J. T., Jr., White, H. J. and Marshall, E. K., Jr. The experimental basis for a method for the quantitative evaluation of the effectiveness of chemotherapeutic agents against streptococcus infection in mice. *J. Pharmacol. and Exper. Therap.*, 67: 437-453, 1939.
11. Litchfield, J. T., Jr. and Fertig, J. W. On a graphic solution of the dosage effect curve. *Bull. Johns Hopkins Hosp.*, 69: 276-286, 1941.

12. MARSHALL, E. K., JR., LITCHFIELD, J. T., JR. and WHITE, H. J. The comparative therapeutic activity of sulfanilamide, sulfapyridine, and diaminosulfone in streptococcus infections in mice. *J. Pharmacol. and Exper. Therap.*, 69: 89-102, 1940.

13. REED, L. J. and MUENCH, H. A simple method of estimating fifty per cent endpoints. *Am. J. Hyg.*, 27: 493-497, 1938.

14. TREVAN, J. W. The error of determination of toxicity. *Proc. Roy. Soc. B*, 101: 483-514, 1927.

15. TREVAN, J. W. A statistical note on the testing of anti-dysentery sera. *J. Path. and Bact.*, 32: 127-134, 1929.

16. WILSON, E. B. and WORCESTER, J. The determination of LD 50 and its sampling error in bio-assay. *Proc. Nat. Acad. Sci., Wash.*, 29: 79-85, 1943.

17. WILSON, E. B. and WORCESTER, J. The determination of LD 50 and its sampling error in bio-assay. II. *Proc. Nat. Acad. Sci., Wash.*, 29: 114-120, 1943.

18. WINSOR, C. P. A comparison of certain symmetrical growth curves. *J. Wash. Acad. Sci.*, 22: 73-81, 1932.

APPENDIX

Observed experiments on the effects of 7 different sulfa drugs in the presence of an infection in mice *

Doses are in mg. per mouse per day unless otherwise specified.

S. A. = Sulfanilamide
D. S. = Diamino-diphenyl-sulphone
S. G. = Sulfaguanadine
E. S. = Ethanol sulfanilamide
S. P. = Sulfapyridine
S. T. = Sulfathiazole
N. P. = Neoprontosil

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
1	S. A.	6.70	0.826	0	4
		4.75	0.677	0	4
		3.88	0.589	1	3
		3.28	0.516	1	3
		2.62	0.418	2	2
		1.95	0.290	3	1
		1.58	0.199	4	0
2	S. A.	4.25	0.628	1	4
		3.42	0.534	1	3
		2.10	0.322	1	3
		1.82	0.260	3	1
		1.41	0.149	2	2
		0.77	$\bar{1}$.886	4	0
		0.58	$\bar{1}$.763	5	0

* This material was kindly furnished to us by Dr. E. K. Marshall, Jr.

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
3	S. A.	8.97	0.953	0	3
		8.23	0.915	1	5
		4.90	0.690	0	5
		4.06	0.609	1	4
		2.60	0.415	2	3
		2.07	0.316	2	1
		1.19	0.279	3	0
4	S. A.	6.72	0.827	0	6
		4.47	0.650	0	6
		3.12	0.494	1	4
		2.58	0.412	2	2
		1.66	0.220	3	2
		1.25	0.097	4	0
5	S. P.	3.82	0.582	0	5
		3.10	0.491	1	4
		1.92	0.283	2	3
		1.62	0.210	3	2
		0.91	$\bar{1}$.959	4	1
		0.66	$\bar{1}$.820	3	2
6	S. P.	1.91	0.281	3	3
		1.28	0.107	6	0
		0.68	$\bar{1}$.833	6	0
		0.37	$\bar{1}$.568	6	0
7	S. P.	4.40	0.643	1	5
		3.20	0.505	1	5
		2.12	0.326	2	4
		1.40	0.146	4	2
		0.97	$\bar{1}$.987	6	0
8	S. P.	3.93	0.594	0	3
		3.60	0.556	0	3
		3.33	0.522	0	3
		2.17	0.336	2	1
		1.77	0.248	2	1
		1.60	0.204	2	1
		1.13	0.053	3	0
		0.80	$\bar{1}$.903	3	0
		0.77	$\bar{1}$.886	3	0
		0.53	$\bar{1}$.724	3	0

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
9	D. S.	0.554	$\bar{1}$.744	1	4
		0.408	$\bar{1}$.611	2	3
		0.278	$\bar{1}$.444	2	3
		0.182	$\bar{1}$.260	5	0
		0.130	$\bar{1}$.114	4	1
		0.082	$\bar{2}$.914	5	0
10	D. S.	0.655	$\bar{1}$.816	0	4
		0.526	$\bar{1}$.721	3	2
		0.332	$\bar{1}$.521	2	3
		0.258	$\bar{1}$.412	4	1
		0.138	$\bar{1}$.140	4	1
		0.068	$\bar{2}$.829	4	0
11	D. S.	0.813	$\bar{1}$.910	0	3
		0.728	$\bar{1}$.862	0	4
		0.540	1.732	2	3
		0.332	$\bar{1}$.521	3	2
		0.179	$\bar{1}$.253	6	3
		0.113	$\bar{1}$.053	3	0
12	D. S.	0.562	$\bar{1}$.750	2	2
		0.378	$\bar{1}$.577	4	4
		0.248	$\bar{1}$.394	3	1
		0.142	$\bar{1}$.152	3	1
		0.110	$\bar{1}$.041	4	0
		0.060	$\bar{2}$.778	4	0
13	S. A.	8.414	0.925	1	9
		6.607	0.820	0	10
		4.954	0.695	1	9
		4.178	0.621	2	8
		3.631	0.560	0	10
		3.251	0.512	3	7
		2.761	0.441	4	6
		2.178	0.338	5	5
		1.901	0.279	7	3
		1.730	0.238	8	2
		1.230	0.090	8	2
		0.5297	1.724	10	0

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
14	S. P.	4.220	0.625	1	11
		3.550	0.550	1	11
		2.560	0.408	3	9
		1.980	0.297	7	5
		1.800	0.255	6	6
		1.430	0.155	10	2
		0.962	$\bar{1}$.983	10	2
		0.788	$\bar{1}$.897	12	0
		0.623	$\bar{1}$.794	10	2
		0.293	$\bar{1}$.467	11	0
15	D. S.	0.723	$\bar{1}$.859	1	11
		0.577	$\bar{1}$.761	4	8
		0.456	$\bar{1}$.659	8	4
		0.345	$\bar{1}$.538	8	4
		0.292	$\bar{1}$.465	3	9
		0.228	$\bar{1}$.358	10	2
		0.161	$\bar{1}$.207	9	3
		0.132	$\bar{1}$.121	10	2
		0.0967	$\bar{2}$.985	12	0
		0.0488	$\bar{2}$.688	8	0
16	S. A.	14.52	1.162	0	5
		8.46	0.927	3	2
		6.04	0.781	3	2
		4.16	0.619	3	2
		2.68	0.428	3	1
		1.10	0.041	4	0
17	S. A.	13.36	1.126	0	5
		8.24	0.916	1	4
		6.48	0.812	2	3
		3.48	0.542	4	1
		1.58	0.199	5	0
		0.60	$\bar{1}$.778	4	0
18	S. A.	12.66	1.102	0	5
		6.84	0.835	3	2
		3.80	0.580	5	0
		1.88	0.274	5	0
		1.31	0.117	8	0

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
19	S. A.	14.52	1.162	0	5
		10.32	1.014	1	4
		8.06	0.906	3	2
		5.70	0.756	2	3
		1.80	0.255	4	0
		0.90	$\bar{1}$.954	4	0
20	S. P.	6.50	0.813	2	3
		3.96	0.598	5	0
		2.90	0.462	3	2
		1.86	0.270	4	1
		0.56	$\bar{1}$.748	5	0
		0.15	$\bar{1}$.176	4	0
21	S. P.	6.90	0.839	3	2
		3.82	0.582	2	3
		1.65	0.217	6	0
		0.53	$\bar{1}$.724	6	0
		0.23	$\bar{1}$.362	6	0
22	S. P.	6.40	0.806	3	2
		5.32	0.726	2	3
		2.96	0.471	4	1
		1.27	0.104	6	0
		0.48	$\bar{1}$.681	6	0
23	S. P.	7.62	0.882	1	4
		6.00	0.778	2	3
		4.14	0.617	2	3
		2.98	0.474	3	2
		1.56	0.193	4	1
		0.95	$\bar{1}$.978	4	0
24	D. S.	1.172	0.069	0	5
		1.018	0.008	1	4
		0.480	$\bar{1}$.681	5	0
		0.308	$\bar{1}$.489	2	3
		0.116	$\bar{1}$.064	5	0
		0.046	$\bar{2}$.663	5	0

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
25	D. S.	1.194	0.077	1	4
		0.834	$\bar{1}$.921	3	2
		0.442	$\bar{1}$.645	2	3
		0.210	$\bar{1}$.322	4	1
		0.104	$\bar{1}$.017	5	0
		0.050	$\bar{2}$.699	5	0
26	D. S.	1.064	0.027	1	4
		0.928	$\bar{1}$.968	1	4
		0.726	$\bar{1}$.861	5	0
		0.474	$\bar{1}$.676	2	3
		0.178	$\bar{1}$.250	4	1
		0.042	$\bar{2}$.623	5	0
27	D. S.	1.360	0.134	2	3
		1.112	0.046	1	4
		0.946	$\bar{1}$.976	2	3
		0.740	$\bar{1}$.869	2	3
		0.282	$\bar{1}$.450	5	0
		0.108	$\bar{1}$.033	5	0
28	S. A.	15.03	1.177	0	12
		11.63	1.066	1	11
		8.58	0.933	5	7
		7.19	0.857	4	8
		5.66	0.753	8	4
		4.02	0.604	10	2
		2.41	0.382	10	2
		1.64	0.215	13	0
		0.86	$\bar{1}$.934	16	0
29	S. P.	7.39	0.869	4	8
		6.08	0.784	6	6
		4.99	0.698	6	6
		3.48	0.542	8	4
		2.91	0.464	8	4
		1.92	0.283	10	2
		1.10	0.041	12	0
		0.61	$\bar{1}$.785	12	0
		0.28	$\bar{1}$.447	17	0

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
30	D. S.	1.300	0.114	3	9
		1.080	0.033	2	10
		0.986	$\bar{1}$.994	3	9
		0.877	$\bar{1}$.943	7	5
		0.612	$\bar{1}$.787	6	6
		0.428	$\bar{1}$.631	9	3
		0.198	$\bar{1}$.297	20	5
		0.057	$\bar{2}$.756	23	0
31	S. A.	45.21*	1.655	3	5
		36.86	1.566	6	2
		24.05	1.381	7	1
		17.08	1.045	16	0
32	S. P.	24.11*	1.382	0	10
		21.90	1.336	2	6
		13.28	1.123	4	2
		8.06	0.906	7	1
		5.62	0.750	8	0
33	D. S.	2.909*	0.464	0	20
		1.740	0.241	0	1
		1.590	0.201	1	0
		1.530	0.185	1	0
		1.530	0.185	1	0
		1.490	0.173	0	1
		1.290	0.111	1	0
		0.657	$\bar{1}$.818	14	0
34	S. T.	23.12*	1.364	1	7
		19.40	1.288	7	3
		12.24	1.088	9	1
		5.99	0.777	12	0
35	S. P.	36.40**	1.561	1	9
		28.67	1.457	2	8
		21.12	1.325	4	6
		17.06	1.232	7	3
		10.76	1.032	9	1

* Mg. per mouse in 6 days.
** Mg. per mouse in 3 days.

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
36	D. S.	2.615**	0.417	0	10
		2.274	0.357	2	8
		1.983	0.297	4	6
		1.605	0.205	4	6
		0.786	$\bar{1}$.895	10	0
37	S. T.	39.20**	1.593	2	8
		32.86	1.517	2	8
		23.80	1.377	5	5
		17.45	1.242	7	3
		8.35	0.922	9	0
38	S. P.	45.870	1.662	1	9
		34.064	1.532	2	8
		22.940	1.361	2	8
		17.370	1.240	5	5
		9.455	0.976	9	1
39	S. T.	44.49	1.648	1	8
		31.47	1.498	2	7
		22.60	1.354	6	4
		17.76	1.294	6	4
		10.07	1.003	8	3
40	D. S.	3.358	0.526	2	10
		2.933	0.467	2	8
		2.264	0.355	2	6
		1.644	0.216	5	3
		0.889	$\bar{1}$.949	6	3
41	E. S.	170.90	2.233	1	12
		92.90	1.968	4	4
		71.96	1.857	8	1
		42.16	1.625	10	0
		26.88	1.429	8	0
42	S. P.	42.06	1.624	0	10
		28.25	1.451	2	8
		20.47	1.311	1	9
		13.82	1.141	4	6
		8.63	0.936	9	0

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
43	S. G.	41.63	1.619	5	4
		29.64	1.472	5	3
		19.90	1.299	9	1
		12.00	1.079	12	0
		6.03	0.780	10	0
44	S. A.	8.507	0.930	3	7
		7.087	0.850	1	9
		4.375	0.641	3	7
		3.392	0.530	4	6
		1.670	0.223	6	1
45	S. T.	8.826	0.946	2	8
		6.921	0.840	3	7
		4.636	0.666	3	4
		3.797	0.579	7	7
		2.082	0.318	7	2
46	S. G.	22.640	1.355	2	7
		16.700	1.223	6	6
		11.110	1.046	6	5
		8.138	0.911	5	3
		4.374	0.641	8	2
47	S. A.	8.939	0.951	3	6
		6.416	0.807	2	8
		4.042	0.607	7	3
		2.795	0.446	8	2
		1.842	0.265	9	1
		1.081	0.034	8	1
48	N. P.	33.430	1.524	1	9
		23.670	1.374	4	6
		16.030	1.205	9	1
		10.980	1.041	9	1
		7.843	0.894	8	2
		5.108	0.708	6	4
49	S. A.	10.44	1.019	2	8
		5.41	0.733	3	7
		4.52	0.655	3	7
		2.83	0.452	10	0
		2.10	0.322	5	5

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
50	N. P.	42.20	1.625	0	10
		22.30	1.348	0	10
		18.20	1.260	5	5
		10.80	1.033	8	2
		7.25	0.860	7	3
51	S. T.	8.52	0.930	1	9
		4.49	0.652	5	5
		3.72	0.571	6	4
		2.33	0.367	5	5
		1.63	0.212	8	2
52	S. A.	22.15**	1.345	1	5
		15.78	1.198	1	5
		10.47	1.020	0	6
		6.68	0.825	3	3
		4.85	0.686	3	3
53	S. A.	19.33**	1.286	1	5
		11.42	1.058	1	5
		8.23	0.915	3	3
		5.10	0.708	1	4
		4.20	0.623	3	2
54	S. P.	22.18**	1.346	1	5
		17.17	1.235	1	5
		11.42	1.058	2	4
		7.35	0.866	1	5
		5.20	0.716	5	1
55	S. P.	20.92**	1.320	0	5
		17.65	1.247	0	6
		9.93	0.997	2	4
		6.85	0.836	2	4
		4.75	0.677	2	4
56	D. S.	1.922**	0.284	1	5
		1.268	0.103	1	5
		0.800	$\bar{1}$.903	2	4
		0.535	$\bar{1}$.728	4	2
		0.403	$\bar{1}$.605	4	2

** Mg. per mouse in 3 days.

APPENDIX (Continued)

EXPERIMENT	DRUG	DOSE	LOG DOSE	DEAD	LIVING
57	D. S.	1.517**	0.181	1	5
		1.167	0.067	2	4
		0.843	$\bar{1}$.926	4	2
		0.470	$\bar{1}$.672	2	4
		0.330	$\bar{1}$.519	3	2

** Mg. per mouse in 3 days.

STATISTICAL ANALYSIS OF THE A–B–O BLOOD GROUPS

BY W. L. STEVENS

Faculdade de Ciências Económicas e Administrativas, Universidade de São Paulo, Rua Vila Nova, São Paulo, Brazil

SUMMARY

THE A-B-O blood group system provides a good example of the statistical estimation of a *pair* of parameters, namely the proportions of A and B genes in the population. Opportunity is taken to show how the common statistical tests for one parameter can readily be generalised to the case of two parameters. To facilitate the application of these tests, a table is provided of the components of the information matrix. A device is suggested for improving the accuracy of these "large-sample" tests. Finally a rapid graphical method is described for judging the significance of the differences between different series of observations.

1. GENERAL PRINCIPLES

1.1. *Introduction*

The statistical analysis of data of the four blood groups belonging to the A-B-O genetic system offers good examples of methods which are generally applicable in all cases where the frequencies are functions of *two* parameters. There are three genes and the proportions of these in the population constitute the parameters. Since however the sum of the three proportions is necessarily unity, there are only *two* independent parameters. The present discussion should therefore be useful, not only for the purpose indicated by the title, but also for illustrating universal methods.

The statistical tests used for two parameters (and thence for more than two) will be seen as simple generalisations of the uniparametric tests. All the tests described are "large-sample" tests; that is to say, they are only strictly true in the limit when the number of observations

tends to infinity. We must therefore not omit to consider how valid they remain when applied to finite samples. The theory of the tests has been expressed in the simplest possible terms in this paper. A full discussion will be found in Stevens (1944).

1.2. *Limiting distributions of estimates*

If x is an estimate of a parameter whose true value is ξ, then when n, the number of observations, is large, x will tend to be normally distributed around ξ with variance (squared standard error) inversely proportional to n. Writing the deviation of the estimate from the parametric value as

$$\delta x = x - \xi,$$

then the variance of x

$$V = v/n$$

where v is a function of ξ but not of n.

If we now write

$$i = 1/v$$

or

$$ni = 1/V, \tag{1.21}$$

then the density of distribution of the estimate is given by

$$\sqrt{\frac{ni}{2\pi}} \exp\left\{-\frac{ni}{2}(\delta x)^2\right\} \tag{1.22}$$

where exp(...) denotes the exponential function.

The quantity i is called the *information per observation* and ni, the *total information.* The latter might also be termed the *invariance,* since it is the inverse of the variance.

Similarly, if there are two parameters ξ and η and their estimates are x and y respectively, then in the limit x and y will be binormally distributed around ξ and η with

$$\begin{aligned} \text{variance } (x), &\qquad V_{xx} = v_{xx}/n, \\ \text{covariance } (x, y), &\qquad V_{xy} = v_{xy}/n, \\ \text{variance } (y), &\qquad V_{yy} = v_{yy}/n, \end{aligned}$$

where v_{xx}, v_{xy} and v_{yy} are functions of ξ and η but not of n. These functions may conveniently be set out in a symmetrical square matrix.

$$\{v\} = \begin{Bmatrix} v_{xx} & v_{xy} \\ v_{xy} & v_{yy} \end{Bmatrix}.$$

Let us now form the reciprocal of this matrix by the ordinary rules of matrix algebra,

$$\{i\} = \begin{Bmatrix} i_{xx} & i_{xy} \\ i_{xy} & i_{yy} \end{Bmatrix}$$

where

$$i_{xx} = v_{yy}/|\, v\,| \;; \qquad i_{xy} = -\, v_{xy}/|\, v\,| \;; \qquad i_{yy} = v_{xx}/|\, v\,| \;;$$

and

$$|\, v\,| = v_{xx}v_{yy} - v_{xy}^2.$$

The process of inverting the matrix may be represented symbolically by

$$\{i\} = 1/\{v\} \tag{1.23}$$

and the matrix $\{i\}$ is called the *information matrix per observation,* while that obtained by multiplying each component by n is called the *total information matrix.*

Writing the deviations of the estimates from the parametric values as

$$\delta x = x - \xi$$
$$\delta y = y - \eta$$

we have the density of the joint distribution of the estimates given by

$$\frac{n\sqrt{|\, i\,|}}{2\pi} \exp\{-\frac{n}{2}(i_{xx}\delta x^2 + 2i_{xy}\delta x\delta y + i_{yy}\delta y^2)\}. \tag{1.24}$$

By observing the nature of the generalisation from equation (1.21) to (1.23) and from the expression (1.22) to (1.24), we shall easily be able to generalise any uniparametric test to the corresponding biparametric test.

1.3. *Comparison of estimates with parametric values*

If the true value of the parameter ξ is given by hypothesis, we usually test the deviation, $\delta x = x - \xi$, by comparing it with the standard error of the estimate. This however is equivalent to writing

$$ni\delta x^2 = \chi^2 \tag{1.31}$$

where χ^2 is a chi-squared with *one* degree of freedom (i. e., the square of a normal deviate with mean zero and standard deviation unity.)

The obvious generalisation for two parameters is then to write

$$n(i_{xx}\delta x^2 + 2i_{xy}\delta x\delta y + i_{yy}\delta y^2) = \chi^2 \tag{1.32}$$

where χ^2 is now a chi-squared with two degrees of freedom. By referring to a table of the integral of χ^2 we can discover if the estimated values, x and y, differ sufficiently from the hypothetical values to necessitate the rejection of the hypothesis.

1.4. *Probability ellipses*

If there is only one parameter, we can calculate a pair of limits

$$\xi = x \pm f\sqrt{V} \tag{1.41}$$

where f is found from a table of the integral of the normal curve. If P is the area outside the ordinates at $-f$ and $+f$, then the probability is $1 - P$ that the true value of the parameter lies within this interval.

Similarly, in the case of two parameters. if we choose the value of χ^2 for *two* degrees of freedom tabulated under probability P, then the probability is $1 - P$ that the point representing the true value of the parameter will lie within the ellipse whose equation is

$$n(i_{xx}\delta x^2 + 2i_{xy}\delta x\delta y + i_{yy}\delta y^2) = \chi^2. \tag{1.42}$$

If the pair of estimates are plotted on rectangular co-ordinates, it is particularly useful to show one or more of these ellipses in order to give an idea of the errors to which the estimates are subject. The choice of a value for P is somewhat arbitrary; $P = 20\%$ is perhaps a good compromise. The probability is then moderately high (80 per cent) that the true value is within the ellipse. For a still higher degree of confidence, we might choose $P = 5\%$, giving a 95 per cent probability that the ellipse will cover the true value.

The ellipse obtained by putting $\chi^2 = 1$ in equation (1.42) may be called the *standard ellipse* by analogy with the *standard error*. The probability is 39 per cent that the true value will lie within the standard ellipse. The median of a χ^2 with two degrees of freedom is 1.386. The ellipse found by putting $\chi^2 = 1.386$ in equation (1.42) may therefore be called the *probable ellipse* by analogy with the *probable error*. The probability that the point representing the true values of the parameters will lie within the probable ellipse is 50 per cent.

1.5. *Comparison of two sets of data*

If we have two samples drawn from the same population, consisting of respectively n_1 and n_2 observations and yielding estimates x_1 and x_2

of a parameter ξ, then the variance of the difference between the estimates, $\delta x = x_1 - x_2$, is

$$\text{variance } (\delta x) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) v.$$

The difference between the estimates can be compared with the standard error of this difference. This however is equivalent to writing

$$\frac{n_1 n_2}{n_1 + n_2}\, i\delta x^2 = \chi^2 \qquad (1.51)$$

where χ^2 is a chi-squared with *one* degree of freedom.

The generalisation for the case of two parameters is then obvious. If x_1 and y_1 are the estimates of ξ and η furnished by the first sample and x_2 and y_2 the estimates furnished by the second sample, and if

$$\delta x = x_1 - x_2 \qquad \delta y = y_1 - y_2$$

then

$$\frac{n_1 n_2}{n_1 + n_2}\, (i_{xx}\delta x^2 + 2i_{xy}\delta x \delta y + i_{yy}\delta y^2) = \chi^2 \qquad (1.52)$$

where χ^2 has *two* degrees of freedom. A significantly large value of χ^2 would indicate that the two samples cannot be considered as having been drawn from the same population.

1.6. *Heterogeneity of more than two sets of data*

Generalising again, we may suppose that we have t (> 2) sets of data and we wish to test whether all these sets might have been drawn from the same population. Let the numbers of observations and the estimates of ξ be respectively $n_1, n_2 \cdots n_t$ and $x_1, x_2 \cdots x_t$ and let

$$N = S(n) = n_1 + n_2 + \cdots + n_t,$$

the weighted total,

$$X = S(nx) = n_1 x_1 + n_2 x_2 + \cdots + n_t x_t,$$

and the weighted mean,

$$\bar{x} = X/N.$$

Then

$$iS\{n(x - \bar{x})^2\} = \chi^2 \qquad (1.61)$$

with number of degrees of freedom $= t - 1$, i. e., one less than the number of sets.

Passing now to the case of two parameters, we have the weighted total,

$$Y = n_1 y_1 + n_2 y_2 + \cdots + n_t y_t,$$

the weighted mean,

$$\bar{y} = Y/N,$$

and the weighted sums of squares and products,

$$S_{xx} = S\{n(x - \bar{x})^2\}; \; S_{xy} = S\{n(x - \bar{x})(y - \bar{y})\}; \; S_{yy} = S\{n(y - \bar{y})^2\}.$$

Then

$$i_{xx} S_{xx} + 2 i_{xy} S_{xy} + i_{yy} S_{yy} = \chi^2 \qquad (1.62)$$

with number of degrees of freedom $= 2(t - 1)$.

A significantly large value of χ^2 will indicate that the samples were not all drawn from the same population. If there is no evidence for heterogeneity, the data may be pooled, in which case the weighted means furnish the general estimates of ξ and η.

1.7. *Validity in finite samples*

There are two reason why the tests described will fail to give exact results when applied to finite samples of data:

(i) the distributions of the estimates are not exactly normal; and

(ii) the components of the information matrix are functions of the parameters. Since we usually do not know the values of the parameters, we are obliged to use the approximate value of the matrix obtained when we insert the respective estimates.

Now it is evident that if, instead of the parameters ξ and η, we use two functions of them, $f(\xi, \eta)$ and $g(\xi, \eta)$, the corresponding estimates will be $f(x, y)$ and $g(x, y)$. In infinitely large samples, any tests of significance involving the transformed parameters and estimates must be exactly equivalent to the corresponding tests made on the original parameters and estimates. In finite samples the results will not be quite the same. We may however be able so to choose the functional transformation that, in finite samples:

(i) the distribution of the transformed estimates will be more nearly normal than that of the original estimates; and

(ii) the components of the information matrix become stabilised, i. e., they change only very slowly with changes in the values of the parameters.

If we can satisfy these two conditions, the tests of significance made on the transformed estimates will be more reliable than those made on the original estimates.

It is not obvious that we shall be able to satisfy these two conditions at the same time, but fortunately in practice it often turns out that in attempting to satisfy one we shall go some way towards satisfying the other. Familiar examples are the logarithmic and square root transformations which are used in preparing certain kinds of data for the analysis of variance. When applicable, these transformations both stabilise the variance and make the distributions more normal.

We therefore suggest that, when one is obliged to use large-sample tests, it is worthwhile to spend a little time on trying to discover some simple transformation which will improve the accuracy of the tests.

2. ESTIMATION OF THE GENETIC FREQUENCIES

2.1. *Efficient estimates*

An estimate of a single parameter is said to be efficient if its standard error is not greater than the standard error of the estimate found by any other method. Similarly for two parameters, a pair of estimates is said to be efficient if no part of the standard ellipse lies outside the standard ellipse of any other method of estimation.

2.2. *Bernstein's method*

Efficient estimates can always be found by the *method of maximal likelihood* (Fisher, 1922). For estimating the proportions of A and B genes, Bernstein (1930) however proposed a method which, although simpler than the method of maximal likelihood, is still fully efficient. Denote the phenotypes and their absolute frequencies by O, A, B, AB and their total by n. Then preliminary (inefficient) estimates of the proportions of the genes A, B and O are given respectively by

$$p' = 1 - \sqrt{\{(O+B)/n\}},$$
$$q' = 1 - \sqrt{\{(O+A)/n\}},$$
$$r' = \sqrt{(O/n)}.$$

The sum of these estimates, instead of equalling unity, falls short by an amount

$$D = 1 - p' - q' - r'.$$

Efficient estimates are now found if we correct the preliminary estimates according to the scheme:

$$p = p'(1 + D/2),$$
$$q = q'(1 + D/2),$$
$$r = (r' + D/2)(1 + D/2).$$

The sum of these three estimates is very nearly exactly unity, thus providing a check on the arithmetic.

Moreover the magnitude of D is a criterion of the agreement of the observed phenotypic frequencies with those given by the genetic hypothesis. The reciprocal of the variance of D is

$$\text{invariance } (D) = 2n(1 + r/pq). \tag{2.23}$$

Hence

$$2n(1 + r/pq)D^2 = \chi^2, \tag{2.24}$$

with *one* degree of freedom. Note that is is sufficient to use the preliminary estimates in this formula.

2.3. *The information matrix*

The general method for finding the information matrix of efficient estimates was given by Fisher (1922) and applied by Stevens (1938) to the problem of the estimation of the gene proportions. Subsequently however Stevens (1944) found explicit formulae for the components of the information matrix, namely

$$i_{pp} = k_{pp}/p \qquad i_{pq} = k_{pq} \qquad i_{qq} = k_{qq}/q \tag{2.31}$$

where [1]

$$k_{pp} = \frac{2(4 - 4p - 6q + 6pq + 2q^2 - 3pq^2)}{(2 - p - 2q)(2 - 2p - q)},$$
$$k_{pq} = \frac{2(4 - 4p - 4q + 3pq)}{(2 - p - 2q)(2 - 2p - q)},$$
$$k_{qq} = \frac{2(4 - 6p - 4q + 2p^2 + 6pq - 3p^2q)}{(2 - p - 2q)(2 - 2p - q)}. \tag{2.32}$$

It will be noticed that, when p and q are small, k_{pp}, k_{pq} and k_{qq} all tend to the value *two*. Hence

$$i_{pp} \to 2/p \qquad \text{and} \qquad i_{qq} \to 2/q.$$

This means that i_{pp} and i_{qq} rise rapidly to infinity near the boundaries,

[1] Strictly speaking, if we retain p, q and r as symbols for the estimates of the gene proportions, we should introduce and use in these formulae other symbols to represent the true parametric values. However as the rest of the paper is concerned only with practical matters, we have not adopted this refinement of notation.

$p = 0$ and $q = 0$. This would make a compact tabulation of these functions quite impossible. We therefore provide instead, in Table 1, the functions k_{pp}, k_{pq} and k_{qq}, from which the components of the information matrix can readily be obtained through equations (2. 31). This table is a condensed version of a three decimal table computed by the author at intervals of 0.01 in p and q. The most useful portions of this more detailed table will be found in Stevens (1945).

TABLE 1

Information matrix

k_{pp}

p \ q	0	0.1	0.2	0.3	0.4
0	2.00	2.00	2.00	2.00	2.00
0.1	2.11	2.13	2.16	2.19	2.22
0.2	2.22	2.28	2.34	2.42	
0.3	2.35	2.45	2.57		
0.4	2.50	2.66			

k_{pq}

p \ q	0	0.1	0.2	0.3	0.4
0	2.00	2.11	2.22	2.35	2.50
0.1	2.11	2.24	2.38	2.55	2.75
0.2	2.22	2.38	2.57	2.79	
0.3	2.35	2.55	2.79		
0.4	2.50	2.75			

Note: The information matrix per observation is

$$\begin{Bmatrix} i_{pp} & i_{pq} \\ i_{pq} & i_{qq} \end{Bmatrix}$$

where

$$i_{pp} = k_{pp}/p$$
$$i_{pq} = k_{pq}$$
$$i_{qq} = k_{qq}/q.$$

k_{qq}

p \ q	0	0.1	0.2	0.3	0.4
0	2.00	2.11	2.22	2.35	2.50
0.1	2.00	2.13	2.28	2.45	2.66
0.2	2.00	2.16	2.34	2.57	
0.3	2.00	2.19	2.42		
0.4	2.00	2.22			

The table may be interpolated linearly with respect to p and q. Below is illustrated the method of double linear interpolation most suitable for a calculating machine, although indeed, after a little practice, these interpolations can be made at sight.

2.4. *Example of the use of the table*

Given $p = 0.1842$ and $q = 0.0682$, find the information matrix. Interpolation of k_{pp} is made according to the following scheme:

p \ q	0	0.068	0.1	Multipliers
0.1	2.11		2.13	16
0.184	2.20	2.24	2.26	
0.2	2.22		2.28	84
Multipliers	32		68	

The four corner values (2.11, 2.13, 2.22 and 2.28) are first copied down from the table of k_{pp}. Next we find the successive differences of p. (The values of p and q may be rounded at the third decimal, and the decimal point may then be dropped.)

p	Δp
0.100	
	84
0.184	
	16
0.200	

The two differences are now *interchanged* and written down in the last column of the scheme. They are then in position for multiplying the tabular values. Thus, interpolating with reference to p, we have

$$(.16)(2.11) + (.84)(2.22) = 2.20;$$
$$(.16)(2.13) + (.84)(2.28) = 2.26.$$

Finally we make a linear interpolation of these two values, with reference to q. The successive differences of q are

$q = 0.000$		0.068		0.100
$\Delta q =$	68		32	.

The differences are *interchanged* and written down in the last line of the scheme. Then,

$$(.32)(2.20) + (.68)(2.26) = 2.24.$$

To find i_{pp}, we have

$$i_{pp} = k_{pp}/p = 2.24/0.1842 = 12.16.$$

The reader will verify that the other two components of the information matrix are

$$i_{pq} = k_{pq} = 2.31;$$
$$i_{qq} = k_{qq}/q = 2.11/0.0682 = 30.94.$$

The true values, for comparison, are 12.142, 2.307 and 30.808. The errors in the values obtained from the table are of no practical consequence.

2.5. *The square-root transformation*

It can be shown that the information matrix of $\sqrt{p}$ and $\sqrt{q}$ is given by

$$\{j\} = \begin{Bmatrix} j_{pp} & j_{pq} \\ j_{pq} & j_{qq} \end{Bmatrix}$$

where

$$j_{pp} = 4k_{pp}; \qquad j_{pq} = 4k_{pq}\sqrt{(pq)}; \qquad j_{qq} = 4k_{qq}.$$

Reference to the table shows that k_{pp}, k_{pq} and k_{qq} change very slowly with p and q. Hence j_{pp} and j_{qq} are both very stable and, while j_{pq} is less stable, it is usually fairly small. This implies that the transformation here proposed—that of taking the square root of p and q—has the second of the properties discussed in section (1. 7) and therefore probably also has the first.

Hence we may expect that tests of significance based on the square roots of the gene proportions will be more reliable than those based directly on the gene proportions. Examples of the two methods will be given later.

2.6. *Equivalent circles*

Although statisticians light-heartedly propose the drawing of probability ellipses, the labour required is considerable and almost prohibitive when a large number of pairs of estimates are to be plotted on the same chart. It is noticeable, for example, that although Haldane (1940) describes how to construct the ellipses and plots on trilinear co-ordinates the estimates from 60 or 70 sets of data, he shows on his chart only one ellipse as an example. Any method for simplifying the problem of constructing the probability ellipses would therefore be extremely welcome.

Now if we use the square roots, $\sqrt{p}$ and $\sqrt{q}$, as rectangular co-ordinates, it will be found that the probability ellipses never depart very much from circles over the region for which the table is provided. When $p = q = 0$, they are in fact true circles. Other ellipses, reduced

to the same area for purposes of comparison, are shown in Fig. 1, for (i) $p = q = 0.2$, (ii) $p = 0$ and $q = 0.4$, and (iii) $p = 0.4$ and $q = 0$. We conclude that no serious error will be introduced if we draw, instead of an ellipse, a circle with the same centre and the same area. Such a circle we shall call the *equivalent circle*.

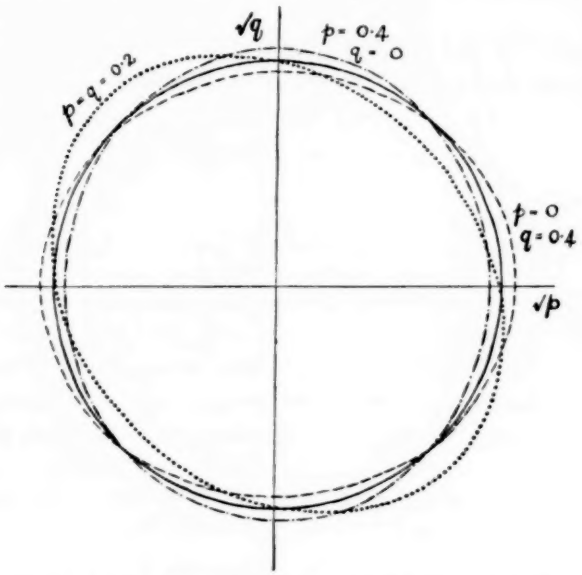

FIG. 1. PROBABILITY ELLIPSES OF THE SQUARE ROOTS FOR REPRESENTATIVE PAIRS OF VALUES OF p AND q, REDUCED TO THE SAME AREA FOR COMPARISON

The radius of the equivalent standard circle may be found from Table 2. The tabular value is to be divided by $\sqrt{n}$, where n is the number of observations.

Example Find the radius of the equivalent standard circle for 250 observations and $p = 0.2232$, $q = 0.1219$. Double linear interpolation in Table 2 gives 0.3348. The radius of the standard circle is therefore

$$\frac{0.3348}{\sqrt{250}} = 0.02117.$$

Of more practical interest however is to find the equivalent circle containing a given percentage, say 80 per cent of the probability. Since

the radius is principally a function of $p + q$ (or of r), the tabulation may be still further simplified by using only $p + q$ (or r) as the argument.

Example Find the radius of the 80 per cent circle for the data of the previous example. We have

$$p + q = 0.345.$$

Interpolating at sight in the column of Table 2 headed 80 per cent, we find,

$$\text{radius} = \frac{0.602}{\sqrt{250}} = 0.0381.$$

TABLE 2

Radius of equivalent circle

STANDARD CIRCLE				CIRCLES INCLUDING $(1-P)$% PROBABILITY					
p or q	0	0.1	0.2	$p+q$	r	50%	80%	95%	99%
0.0	0.3536	0.3490	0.3444	0.0	1.0	0.416	0.634	0.865	1.073
.1	.3490	.3436	.3378	.1	0.9	.411	.626	.854	1.059
.2	.3444	.3378	.3307	.2	.8	.405	.617	.842	1.044
.3	.3395	.3316	.3230	.3	.7	.399	.607	.829	1.027
.4	.3344	.3249	.3144	.4	.6	.391	.596	.813	1.007
.5	.3290	.3175		.5	.5	.381	.581	.793	0.983

Note: To find the radius divide the tabular value by the square root of the number of observations.

We might as well have dispensed with interpolation and merely used the nearest tabular entry. The difference would not be perceptible on a chart.

2.7. *Graphical representation of results*

Trilinear co-ordinates have often been favoured for representing the gene proportions. The reason is not very clear: trilinear co-ordinates are essentially useful for representing non-metrical properties. The discussion of the previous section suggests how we might set up a co-ordinate system with some very useful properties. We use rectangular co-ordinates of $\sqrt{p}$ and $\sqrt{q}$ but label them in terms of p and q as in Fig. 2. It will be noticed that the contours of r are quadrants of a circle. A scale, prepared from Table 2, shows the radius of the 80 per cent circle as a function of r and n.

Given a series of pairs of estimates, p and q, we may plot them directly on this grid, drawing round each point a circle with radius

taken directly from the scale. Then we shall be able to see at a glance which sets of observations differ significantly and which might have been drawn from the same population. If two circles overlap, the hypothesis that the two sets were drawn at random from the same population is acceptable and there is therefore no *statistical* objection to pooling the two sets of data. If two circles do not overlap, the sets

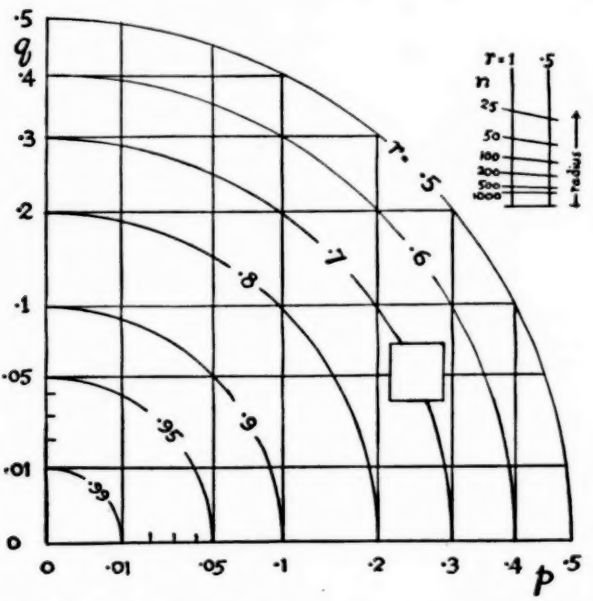

FIG. 2. A CO-ORDINATE GRID WHICH TRANSFORMS ALL PROBABILITY ELLIPSES APPROXIMATELY TO CIRCLES
The radius of the 80 per cent circle is given by the scale at the side.

of data came almost certainly from two different populations (or, alternatively, the two sets, if known to be from the same population, cannot have been random samples.) Calculated tests of significance become unnecessary except in border-line cases or for testing the overall heterogeneity of many sets of data for which some of the circles overlap and some do not.

Of course, the proposed grid of co-ordinates could be much improved by drawing to a larger scale and omitting unnecessary regions. Fig. 2 is only intended to show the principle of construction.

3. APPLICATIONS

3.1. *Estimation from a single series*

The data used for the example are a London series of 422 observations made by Taylor and Prior. We follow Bernstein's method as described in section (2.2).

O	*A*	*B*	*AB*	*n*
202	179	35	6	422

Preliminary estimates:

$$\sqrt{\{(O+B)/n\}} = \sqrt{0.56161} = 0.74941, \qquad p' = 0.25059$$
$$\sqrt{\{(O+A)/n\}} = \sqrt{0.90284} = 0.95017 \qquad q' = 0.04983$$
$$\sqrt{\{(O/n\}} = \sqrt{0.47867} = 0.69186 \qquad r' = 0.69186$$
$$\text{Total} = 0.99228$$
$$D^2 = 0.519 \times 10^{-4} \qquad D = 0.00772$$

Test of agreement with genetic hypothesis:

$$\text{Invariance } (D) = 2n(1 + r/pq) = (844)(56.4) = 4.76 \times 10^4$$
$$\chi^2 = (0.519)(4.76) = 2.42 \text{ (not significant).}$$

Revised estimates:

$$1 + D/2 = 1.00386 \qquad r' + D/2 = 0.69572$$
$$p = p'(1 + D/2) = 0.25156$$
$$q = q'(1 + D/2) = 0.05002$$
$$r = (r' + D/2)(1 + D/2) = 0.69841$$
$$p + q + r = 0.99999 \text{ (checks)}$$

Information matrix: Table 1 gives us

$$\begin{Bmatrix} k_{pp} & k_{pq} \\ & k_{qq} \end{Bmatrix} = \begin{Bmatrix} 2.33 & 2.38 \\ & 2.09 \end{Bmatrix}.$$

The total information matrix is therefore

$$\begin{Bmatrix} ni_{pp} & ni_{pq} \\ & ni_{qq} \end{Bmatrix} = \begin{Bmatrix} nk_{pp}/p & nk_{pq} \\ & nk_{qq}/q \end{Bmatrix} = \begin{Bmatrix} 3909 & 1004 \\ & 17633 \end{Bmatrix}.$$

The value of χ^2 for two degrees of freedom and $P = 20\%$ is 3.2189. The equation of the ellipse containing 80 per cent of the probability is therefore:

$$3907(p - 0.25156)^2 + 2008(p - 0.25156)(q - 0.05002) + 17633(q - 0.05002)^2 = 3.22.$$

This ellipse is shown in Fig. 3a.

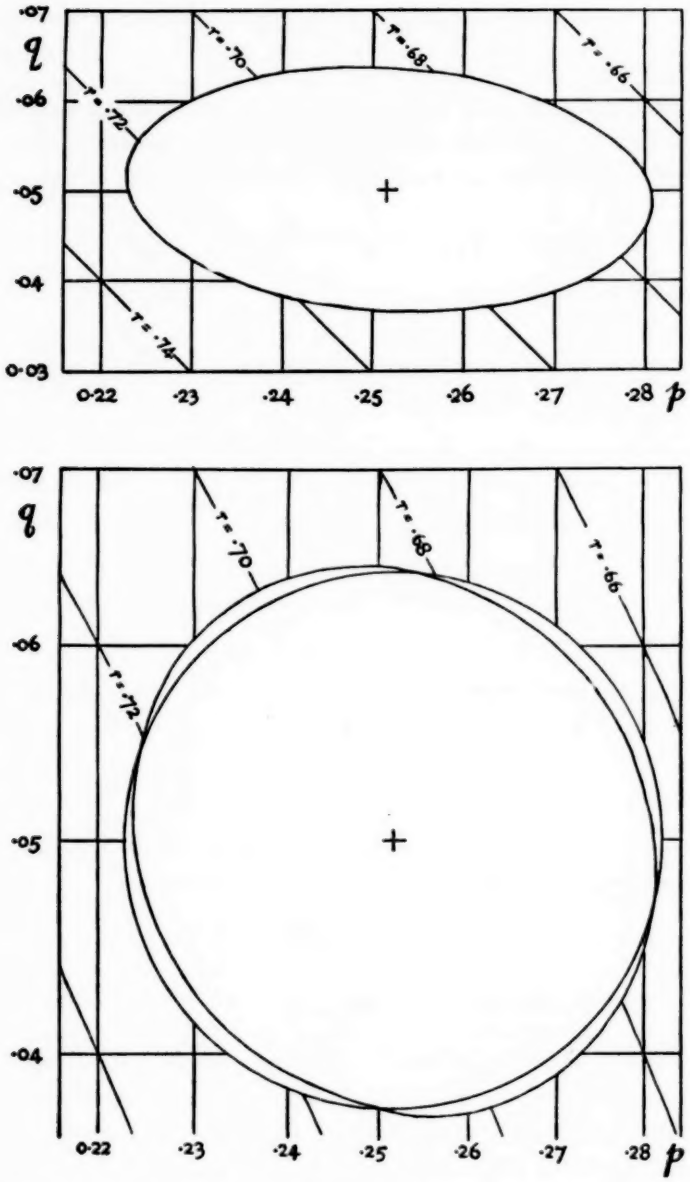

FIG. 3a (ABOVE). THE 80% PROBABILITY ELLIPSE FOR THE DATA IN SECTION (3.1)

FIG. 3b (BELOW). THE 80% PROBABILITY ELLIPSE AND THE EQUIVALENT 80% CIRCLE AFTER A SQUARE ROOT TRANSFORMATION OF THE ESTIMATES

3.11. *Square root transformation*

Alternatively we may calculate the square roots of the estimates:

$$\sqrt{p} = 0.50156;$$
$$\sqrt{q} = 0.22365; \qquad \sqrt{(pq)} = 0.11217.$$

The total information matrix is then

$$\begin{Bmatrix} nj_{pp} & nj_{pq} \\ & nj_{qq} \end{Bmatrix} = \begin{Bmatrix} 4nk_{pp} & 4n\sqrt{(pq)} \cdot k_{pq} \\ & 4nk_{qq} \end{Bmatrix} = \begin{Bmatrix} 3933 & 451 \\ & 3528 \end{Bmatrix}$$

and the equation of the ellipse containing 80 per cent of the probability is

$$3933(\sqrt{p} - 0.50156)^2 + 902(\sqrt{p} - 0.50156)(\sqrt{q} - 0.22365) + 3528(\sqrt{q} - 0.22365)^2 = 3.22.$$

Finally we may calculate the radius of the equivalent circle. Entering Table 2 with $r = 0.7$ in the 80 per cent column, we find:

$$\text{radius} = \frac{0.607}{\sqrt{422}} = 0.0295.$$

Both the ellipse and the circle are shown in Fig. 3b. The original scales are of course in $\sqrt{p}$ and $\sqrt{q}$ but they are re-labelled in terms of p and q. The rectangular grid is in fact a portion extracted from the complete grid of Fig. 2, the place from whence it is taken being there indicated.

3.2. *Comparison with given parametric values*

It can rarely happen in practice that we want to test the agreement of the estimated gene proportions with values given by some hypothesis, but for completeness of exposition we shall imagine such an example.

Suppose then that we wish to use the data of the previous example to test the hypothesis that the true values are

$$p = 0.26444; \qquad q = 0.03578.$$

Although one would not go far wrong in using the information matrix previously calculated, the theoretically correct procedure is to find the matrix which corresponds, not to the estimated values, but to the parametric values given by the hypothesis. Using the above values of p and q, we find

$$\begin{Bmatrix} k_{pp} & k_{pq} \\ & k_{qq} \end{Bmatrix} = \begin{Bmatrix} 2.33 & 2.37 \\ & 2.06 \end{Bmatrix}.$$

The information matrix *per observation* is therefore

$$\begin{Bmatrix} i_{pp} & i_{pq} \\ & i_{qq} \end{Bmatrix} = \begin{Bmatrix} 8.81 & 2.37 \\ & 57.57 \end{Bmatrix}.$$

The deviations from hypothesis are

$$\delta p = 0.25156 - 0.26444 = -0.01288,$$
$$\delta q = 0.05002 - 0.03578 = +0.01424.$$

Hence:

$$\delta p^2 = 1.659; \qquad \delta p \delta q = -1.834; \qquad \delta q^2 = 2.028 \times 10^{-4};$$
$$\chi^2 = n(i_{pp}\delta p^2 + 2i_{pq}\delta p \delta q + i_{qq}\delta q^2)$$
$$= 422\{(8.81)(1.659) - 2(2.37)(1.834) + (57.57)(2.028)\}10^{-4}$$
$$= 5.18. \qquad \text{(Degrees of freedom} = 2)$$

The deviation from hypothesis is accordingly judged to be not significant, although since χ^2 does pass the 10 per cent level there are some grounds for doubting the truth of the hypothesis put forward.

3.21. *Square root transformation*

The square roots of the values given by hypothesis are

$$\left.\begin{matrix} \sqrt{p} = \sqrt{0.26444} = 0.51424 \\ \sqrt{q} = \sqrt{0.03578} = 0.18916 \end{matrix}\right\} \sqrt{(pq)} = 0.09727.$$

After a square root transformation, the information matrix per observation becomes

$$\begin{Bmatrix} j_{pp} & j_{pq} \\ & j_{qq} \end{Bmatrix} = 4 \begin{Bmatrix} k_{pp} & k_{pq}\sqrt{(pq)} \\ & k_{qq} \end{Bmatrix} = 4 \begin{Bmatrix} 2.33 & 0.23 \\ & 2.06 \end{Bmatrix}.$$

The deviations from hypothesis are

$$\delta\sqrt{p} = 0.50156 - 0.51424 = -0.01268,$$
$$\delta\sqrt{q} = 0.22365 - 0.18916 = +0.03449.$$

Hence:

$$(\delta\sqrt{p})^2 = 1.608; \qquad (\delta\sqrt{p})(\delta\sqrt{q}) = -4.373;$$
$$(\delta\sqrt{q})^2 = 11.896 \times 10^{-4};$$
$$\chi^2 = (422)(4)\{(2.33)(1.608) - 2(0.23)(4.373)$$
$$+ (2.06)(11.896)\}10^{-4}$$
$$= 4.43. \qquad \text{(Degrees of freedom} = 2)$$

3.22. *Discussion*

The difference between the values of χ^2 found by the two methods ($\chi^2 = 5.18$ and $\chi^2 = 4.43$) is appreciable though not serious. There is no doubt that the value found after a square root transformation is the more reliable, although to justify this assertion over a wide range would require an enormous amount of arithmetic verification. The use of the square root transformation is indicated when *either* the total number of observations is small *or* either p or q is small. In any event, the amount of extra computation required is not so great as to discourage the method.

We might also use the graphical method, plotting the estimates and the hypothetical values on the square root grid already described. Let three circles be drawn round the hypothetical point with radii calculated from the last three columns of Table 2. Then according to where the point representing the estimates falls in relation to these circles, we may judge the significance of the deviation. The level of significance is of course the complement of the percentage which we have used to label the circle. Unless the quantity of data is very large, this approximate test of significance will be not less reliable than the first of the tests described.

3.3. *Comparison of two sets of data*

The data to be examined come from a collection made by Haldane (1940):

Series	Origin	Author	*O*	*A*	*B*	*AB*	*n*
I	London, Hammersmith	Vaughan	667	600	120	45	1432
II	London, Central (English)	Taylor	1545	1575	303	113	3536
Total			2212	2175	423	158	4968

The question to be answered is whether these two series may legitimately be regarded as having been drawn at random from the same population.

Estimates of p and q are found by Bernstein's method for both series separately and for the combined data. Agreement with the genetic hypothesis must also be verified as explained in section (3.1). It was found that agreement was satisfactory in both series but the calculations are not shown here.

Series	p	q	r
I	0.25859	0.05935	0.68206
II	0.27720	0.06071	0.66209
Combined data	0.27182	0.06031	0.66787
Weighted means	0.27184	0.06032	
Differences	$-0.01861 = \delta p$	$-0.00136 = \delta q$	

The estimates for the pooled data are also found as weighted means of the individual estimates, using the number of observations as weights,

$$\frac{(1432)(0.25859) + (3536)(0.27720)}{4968} = 0.27184, \text{ etc.}$$

Comparison of these two weighted means with the estimates found from the combined data provides a useful check. Unless the number of observations is very small the agreement should be very close.

The information matrix per observation, for $p = 0.27182$ and $q = 0.06031$, is found to be

$$\begin{Bmatrix} i_{pp} & i_{pq} \\ & i_{qq} \end{Bmatrix} = \begin{Bmatrix} 2.36/p & 2.42 \\ & 2.11/q \end{Bmatrix} = \begin{Bmatrix} 8.68 & 2.42 \\ & 34.99 \end{Bmatrix}.$$

$$\delta p^2 = 3.463; \quad \delta p \delta q = 0.253; \quad \delta q^2 = 0.018 \times 10^{-4};$$
$$n_1 n_2/(n_1 + n_2) = (1432)(3526)/4968 = 1019.2.$$

Hence, by equation (1.52) of section (1.5), we have

$$\chi^2 = (1019.2)\{(8.68)(3.463) + 2(2.42)(0.253) + (34.99)(0.018)\}10^{-4}$$
$$= 3.25. \qquad \text{(Degrees of freedom} = 2)$$

The differences between the estimates from the two series are therefore judged to be not statistically significant and there is no statistical objection to pooling the data and presenting the common estimates which have already been calculated.

3.31. *Square root transformation*

Alternatively we may proceed via the square roots as follows:

	$\sqrt{p}$	$\sqrt{q}$
Series I	0.50852	0.24362
Series II	0.52650	0.24639
Differences	−0.01798	−0.00277

$$(\delta\sqrt{p})^2 = 3.233; \qquad (\delta\sqrt{p})(\delta\sqrt{q}) = 0.498;$$
$$(\delta\sqrt{q})^2 = 0.077 \times 10^{-4};$$
$$\sqrt{(pq)} = \sqrt{\{(0.27182)(0.06031)\}} = 0.1280.$$

The information matrix per observation is

$$\begin{Bmatrix} j_{pp} & j_{pq} \\ & j_{qq} \end{Bmatrix} = 4 \begin{Bmatrix} 2.36 & 2.42\sqrt{(pq)} \\ & 2.11 \end{Bmatrix} = 4 \begin{Bmatrix} 2.36 & 0.31 \\ & 2.11 \end{Bmatrix}.$$

Hence

$$\chi^2 = (1019.2)(4)\{(2.36)(3.233) + 2(0.31)(0.498) + (2.11)(0.077)\}10^{-4}$$
$$= 3.30. \qquad \text{(Degrees of freedom} = 2)$$

This value of χ^2 differs very little from that obtained by the direct method. Both may therefore be considered as quite reliable and the use of the square root transformation in this example was not really necessary.

3.4. *Comparison of several sets of data*

The data to be examined consist of the two series considered in the previous example together with five other series also taken from Haldane's collection.

Series	Origin	Author	*O*	*A*	*B*	*AB*	*n*
	South East England						
I	London, Hammersmith	Vaughan	667	600	120	45	1432
II	London, Central (English)	Taylor	1545	1575	303	113	3536
III	London, Central (Scottish)	Taylor	71	59	24	6	160
IV	Slough	Vaughan	2009	1585	334	104	4032
	Ireland						
V	Northern Ireland	Taylor	207	98	43	6	354
VI	Rachrai	MacConaill	10	11	6	4	31
VII	Dublin	Boyd & Boyd	220	124	48	7	399

The problem is to decide which sets of data (if any) may legitimately be regarded as having been drawn from the same population. The most economical method for solving such a problem is to proceed directly to the graphical construction described in section (2.7). Each series was tested for agreement with genetic hypothesis and found to be satisfactory. Estimates were found by Bernstein's method and their square roots calculated. The values of R, the radii of the equivalent 80 per cent circles were found from Table 2, as explained in section (2.6).

Series	p	q	$\sqrt{p}$	$\sqrt{q}$	R
I	0.25859	0.05935	0.50852	0.24362	0.0160
II	.27720	.06071	.52650	.24639	.0101
III	.23011	.09890	.47965	.31448	.0478
IV	.23776	.05590	.48761	.23643	.0096
V	.15995	.07193	.39994	.26820	.0326
VI	.27784	.17460	.52711	.41785	.1056
VII	.18094	.07167	.42537	.26771	.0306

The estimates and their 80 per cent probability circles are shown in Fig. 4. Over the region covered by this chart, the radii are almost entirely functions only of n. It is therefore possible to give a general scale for the radius at the side of the chart.

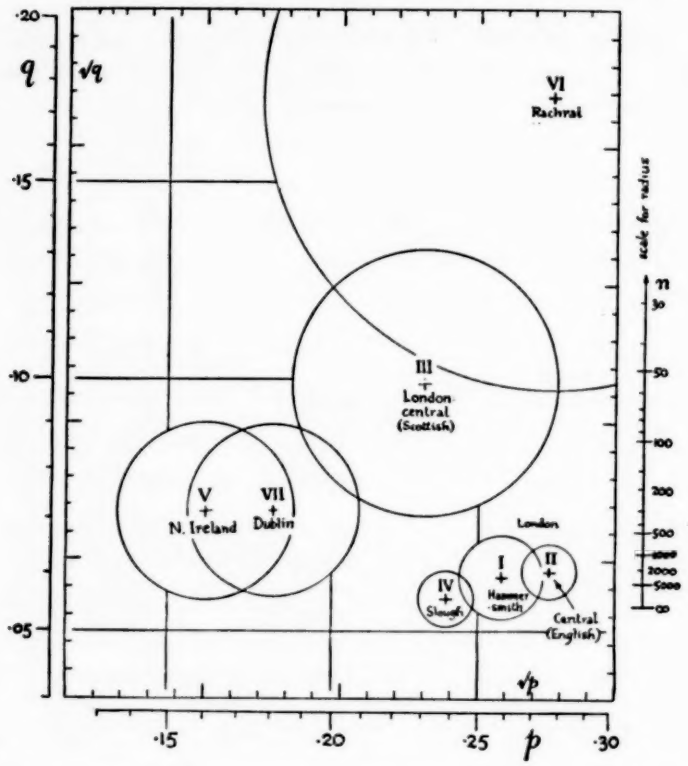

FIG. 4. THE 80% EQUIVALENT CIRCLES FOR THE SEVEN SETS OF DATA EXAMINED IN SECTION (3.4)

A glance at the chart is sufficient to reveal which sets of data may be pooled and which may not. The two main Irish series may be combined without hesitation but the smaller series for Rachrai is certainly from a different population. Scotsmen in London (III) show significantly different gene proportions from the other London series. There is no objection, on statistical grounds, to combining Hammersmith (I) with either Slough (IV) or with London, Central, English (II). But these last three cannot be considered as all having been drawn from the same population; the circles for II and IV are much too far apart.

Computed tests of significance are not likely to tell us anything different, but, in order to illustrate the method, we will test the homogeneity of the four English series.

3.41. *Test of homogeneity of four series*

The subtotals for the four English series (I, II, III and IV) are

O	*A*	*B*	*AB*	*n*
4292	3819	781	268	9160

whence we find $p = 0.25590$ and $q = 0.05902$.

Estimates for the four series separately are collected from the earlier table and multiplied by the respective number of observations. Computations can be reduced by omitting the initial "2" in the values of p.

Series	n	$p(-0.2)$	q	np	nq
I	1432	0.05859	0.05935	83.90	84.99
II	3536	.07720	.06071	272.98	214.67
III	160	.03011	.09890	4.82	15.82
IV	4032	.03776	.05590	152.25	225.39
Total	$9160 = N$			$P = 513.95$	$540.87 = Q$

The weighted means of p and q are

$$\bar{p} = 0.2 + P/N = 0.2 + 513.95/9160 = 0.25611;$$
$$\bar{q} = \quad Q/N \quad = \quad 540.87/9160 \quad = 0.05905.$$

These agree sufficiently closely with the estimates derived from the pooled data, $p = 0.25590$ and $q = 0.05902$. Using the latter values, we find

$$\begin{Bmatrix} k_{pp} & k_{pq} \\ & k_{qq} \end{Bmatrix} = \begin{Bmatrix} 2.34 & 2.40 \\ & 2.11 \end{Bmatrix}.$$

Hence

$$\begin{Bmatrix} i_{pp} & i_{pq} \\ & i_{qq} \end{Bmatrix} = \begin{Bmatrix} 9.14 & 2.40 \\ & 35.75 \end{Bmatrix}.$$

Next we calculate the weighted sums of squares and products, by summing the products of terms in two columns of the table above. Thus for squares of p, we have:

$$S(np^2) = (.05859)(83.90) + (.07720)(272.98) + \cdots, \text{ etc.}$$
$$= 31.884$$
$$P^2/N = (513.95)^2/9160 \quad = 28.837$$
$$\text{Hence} \qquad S_{pp} = S\{n(p-\bar{p})^2\} \quad = \overline{3.047}$$

Similarly, we have

$$S(npq) = 30.540 \qquad S(np^2) = 32.341$$
$$PQ/N = 30.347 \qquad Q^2/N = 31.937$$
$$S_{pq} = 0.193 \qquad S_{qq} = 0.304.$$

Hence

$$\chi^2 = i_{pp}S_{pp} + 2i_{pq}S_{pq} + i_{qq}S_{qq}$$
$$= (9.14)(3.047) + 2(2.40)(0.193) + (35.75)(0.304)$$
$$= 39.64. \qquad [\text{Degrees of freedom} = 2(4-1) = 6]$$

This value of χ^2 is extremely significant, as we had already anticipated.

3.42. *Square root transformation*

Alternatively we apply the same test to the square roots of the estimates instead of to the estimates themselves. The information matrix per observation is

$$\begin{Bmatrix} j_{pp} & j_{pq} \\ & j_{qq} \end{Bmatrix} = 4 \begin{Bmatrix} k_{pp} & k_{pq}\sqrt{(pq)} \\ & k_{qq} \end{Bmatrix} = 4 \begin{Bmatrix} 2.34 & 0.29 \\ & 2.11 \end{Bmatrix}.$$

The weighted sums of squares and products of the square roots are found to be

$$2.969; \qquad 0.433; \qquad 1.032.$$

Hence

$$\chi^2 = 4\{(2.34)(2.969) + 2(0.29)(0.433) + (2.11)(1.032)\}$$
$$= 37.50. \qquad (\text{Degrees of freedom} = 6)$$

This differs appreciably though not seriously from the value found by the direct method. The discrepancy can be attributed principally to the presence of the short series of only 160 observations. Had all the series been of over 1000 observations, the difference between the values of χ^2 calculated by the two methods would have been negligible. The latter value of χ^2 may be considered the more reliable. though of course the conclusion is the same either way.

3.5. *Partition of chi-squared*

The χ^2 which represents the overall heterogeneity of several sets of data may be partitioned into orthogonal components. How we choose to make the separation will of course depend on circumstances. To

illustrate the process with the present example, we may consider the following partition:

Comparison			Degrees of freedom
I *versus* II	:	Hammersmith compared with Central (English)	2
(I + II) *vs.* IV	:	London compared with Slough	2
(I + II + IV) *vs.* III	:	London (Scottish) compared with the rest	2
		Total	6

Each χ^2 with two degrees of freedom may be calculated by the method of section (3. 3). If this procedure is followed, the partition will be inexact, i. e., the three values of χ^2 will add only approximately to the total, $\chi^2 = 39.64$ (or 37.50 if we are using the square roots). Alternatively we may follow the method of section (3. 3) but use always the information matrix which we have found for the combined data of the *four* series. In this case, the three components of χ^2 will add exactly to the right total but each component χ^2 (except the last) will contain a small error. Unless the various pairs of estimates differ very widely (in which case one would not be bothering with a test of significance) the errors introduced will not be of any practical consequence. It may be noted that these errors are smaller if we use the square root transformation because the information matrix of the square roots is considerably more stable.

Following the method of section (3. 3) but using the information matrix of section (3. 41), we find for the first comparison in the above partition

$$\chi^2 = 3.42.$$

This differs but little from the value 3.25 found previously. The computation of the other components is effected most rapidly if we use estimates obtained from the weighted means; when we pool two series, we find the subtotals of n, np and nq, dividing the last two by the first to give p and q for the pooled data.

Series	n	$p\,(-0.2)$	q	np	nq
I + II	4968	0.07184	0.06032	356.88	299.66
IV	4032	.03776	.05590	152.25	225.39
Total	9000			509.13	525.05
Difference		$\delta p = .03408$	$\delta q = .00442$		

Hence, by the method of section (3. 3) but using the matrix of section (3. 41), we find

$$\chi^2 = 26.79.$$

Finally we have

Series	n	$p(-0.2)$	q	np	nq
I + II + IV	9000	0.05657	0.05834	509.13	525.05
III	160	.03011	.09890	4.82	15.82
Total	9160			513.95	540.87
Difference		+ 0.02646	— 0.04056	(Checks)	

giving

$$\chi^2 = 9.44.$$

The partition of χ^2 therefore reads:

Comparison	Degrees of freedom	χ^2	Significance
I *versus* II	2	3.42	not significant
(I + II) *versus* IV	2	26.79	very significant
(I + II + IV) *versus* III	2	9.44	very significant
Total	6	39.64	

It will be observed that the three components sum exactly (but for the rounding error) to the χ^2 of six degrees of freedom which was computed directly in section (3. 41).

3.6. *Procedure when original data are not available*

It sometimes happens that a research worker wishes to collate the results of other authors who publish the estimated gene frequencies but omit the original data. If nothing is known of how these estimates were obtained, none of the tests of significance described in this paper can be used. If on the other hand we know both the number of observations and that the estimates were found by Bernstein's or any other efficient method, then there is no difficulty in applying any of these tests. The only modification is that the weighted means must furnish the estimates for pooled data, instead of being available as checks. Similar remarks will apply in any problem of testing and combining statistical estimates.

REFERENCES

BERNSTEIN, F. (1930). Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen. *Z. indukt. Abstamm.-VererbLehr*, 61, Heft 2, 233.

FISHER. R. A. (1922). On the mathematical foundations of theoretical statistics. *Phil. Trans.*: A, 222, 309-368.

HALDANE. J. B. S. (1940). The blood-group frequencies of European peoples, and racial origins. *Human Biology*, 12, 457-480.

STEVENS, W. L. (1938). Estimation of blood-group gene frequencies. *Ann. Eugenics*, 8, 362-375.

——— (1944). Estimação Estatística. *Revista da Faculdade de Ciências da Universidade de Coimbra (Portugal)*.

——— (1945). Tabelas para investigações sôbre os grupos sanguíneos. *Revista Fac. Ciências Coimbra.*

NOTES

A NOTE ON THE RELATIONSHIP OF BIRTHS TO DAY OF WEEK

IN A SYMPOSIUM celebrating the centennial of the birth of Ribot, there is a paper by Karl Marbe (1) concerning psychology and statistics. Marbe's thesis is the familiar one that some of the errors which occur in statistical research arise from the operation of psychological factors, such as bias, prejudice, popular belief, and the like. Among his illustrative data is a table reporting the average number of births of children in Czechoslovakia by day of week for each of a series of six consecutive years.

Marbe points out a systematic trend in births, Tuesday being the day of least registration, and Sunday being the day of most registration, with a fairly constant trend of the figures in between these dates in the weekly cycle. This trend is apparent in each of the years reported separately and, of course, in the averages (see Table 1). This phe-

TABLE 1

Births by day of week, from three sources

	Czechoslovakia	Minneapolis				Nurnberg
	6 yr. average, 1925-30	Average, 1946	Per cent of total	Average, 9 mos., 1949	Per cent of total	Per cent of total, 1935
Sunday	1006.7	38.1	13.2	44.4	13.5	14.8
Monday	922.0	39.7	13.7	44.2	13.4	13.4
Tuesday	913.1	44.1	15.5	49.9	15.1	14.2
Wednesday	927.2	40.9	14.1	47.9	14.5	13.9
Thursday	967.8	40.8	14.1	47.1	14.3	13.7
Friday	952.3	43.0	14.8	49.6	15.0	14.4
Saturday	977.6	42.0	14.6	46.4	14.1	15.6
			100.0		99.9	100.0

nomenon is considered an illustration of the operation of a popular superstition, presumably relating to folklore about Sunday being a favorable day for children to be born, with the result that births are reported as occurring on Sunday in excess of fact.

The significance of these figures Marbe finds in the fact that Czechoslovakia presumably had (pre-war) the most efficient system of vital statistics on the European continent, and records, therefore, should be relatively more free from the operation of psychological factors than data from other countries. His interpretation is clearly an admissable one, showing the contaminating effect of psychological or attitudinal factors in statistical research.

There is another interpretation of these figures, that they are sound figures and that the trend so found reflects an experiential or cultural factor. Such an hypothesis would hold that, probably due to systematic variation in work or exercise or other activity relating to the cycle of the week, births are precipitated at certain days more frequently than at other periods. For example, Nurnberger (3) quotes U. Winckel's assertion that pregnancies which terminate in summer are slightly shorter than those which terminate in winter. This difference he attributes to the greater exercise taken in pleasant weather.

The only other birth data relating to day of week which this writer could discover appears in an article by Meyer (2), who reports percentages of a total of 6,385 births in the city of Nurnberg for 1935 for each day of the week (see Table 1). In these figures, there is a noticeable lack of trend, although Saturday and Sunday are the most popular days, as in the data reported by Marbe.

To check Marbe's findings, birth data for the city of Minneapolis, the largest birth registration area in the state of Minnesota, for the year 1946 were analyzed according to day of the week. Table 1 shows the averages for each day of the week.

Clearly, there is no systematic trend by day of week. Variance analysis, however, applied to the original observations showed that variation between the means of days and variation between means of weeks are both significantly different from the remainder variance, the F ratios being 4.9 and 13.6 respectively, well beyond the one per cent probability level (see Table 2). While the numbers dealt with are very much smaller than those reported by Marbe, any trend which would be of significance on a large scale would be apparent in data for a large city. Marbe's implicit hypothesis that there is no true systematic variation of births by day of week is, therefore, refuted by these data.

Since 1946 was an unusual year, in that the number of births increased steadily week by week throughout the year, the writer secured further data. Material was available for the first 38 full weeks of 1949, extending from Sunday, January 2 to and including Saturday, September 24. Table 3 gives variance data for this year. Again, the portions of total

TABLE 2

Analysis of variance

Minneapolis birth data for 1946

SOURCE OF VARIATION	SUM OF SQUARES	D. F.	MEAN SQUARE
Total	42 763	363*	
Weeks	28 828	51	565.
Days	1 218	6	203.
Remainder	12 717	306	41.6

* December 31st was dropped for symmetry.

TABLE 3

Analysis of variance

Minneapolis birth data for 1949 (38 weeks)

SOURCE OF VARIATION	SUM OF SQUARES	D. F.	MEAN SQUARE
Total	14 745	265	
Weeks	3 411	37	92.2
Days	1 173	6	195.5
Remainder	10 161	222	45.8

variance attributable to both weeks and days is significantly greater than the remainder variance. In both years, Sunday and Monday are significantly lower than the general average, and Tuesday and Friday are significantly higher.

Probably sufficient births occur under medical control to introduce the significant differences observed. Approximately seven per cent of births at the University of Minnesota hospitals are Caesarean.* Although this figure is probably higher than for hospitals in general, Caesarean births plus forms of induced delivery probably serve to explain the variation noted in Table 1. Sunday and Monday would be low because of

* The writer is indebted to Dr. Charles P. Winsor, Editor of this Journal for this suggestion. The proportion of Caesarean births reported by him as occurring at the Johns Hopkins Hospital is five per cent.

the holiday on the former, and staff conferences on the latter, day. Such explanations appear to interpret the data more adequately than " psychological " or " sociological " hypotheses.

REFERENCES

1. MARBE, KARL. Über psychologie und Statistik. *In*: Centenaire de Th. Ribot. *Agen, France*, 1939, pp. 445-58.
2. MEYER, M. Zur Frage der Geburtzeit der Kinder. *Med. Welt*, 1936, 10: 907-09.
3. NURNBERGER, LUDWIG. Abnormal Schwangerschaftsdauer. *In*: HALBAN, J. and SEITZ, L., Biologie und Pathologie des Weibes. Vol. 7, Part I, p. 374. *Berlin*, 1927.

DALE B. HARRIS,
Institute of Child Welfare,
University of Minnesota.

QUOTATIONS

THE SCIENCE OF ST. GEORGE MIVART

FOR the popular historian, men make history; he sees the historical process as a succession of discrete individuals, the contribution of each providing the period of time during which he lived and worked with the distinctive flavor of his particular abilities. Historical movements and events are evaluated and interpreted, and, in some cases, come alive through the identification of a single figure, as genius, with the social whole of which he was but a part. One speaks of the Napoleonic Era, of the Newtonian or Copernican Age; and in the natural sciences, the name of Darwin comprehends a whole intellectual movement which revolutionized scientific thinking in the latter part of the nineteenth century.

In so labelling such periods by the name of their dominant figures, it is often possible to lose sight of the multitude of less luminous lights all of whom can be said to have contributed, in proportion to their abilities, to the popular acceptance of the movement of which they often remain anonymous members. Darwin, with perhaps a subscript Huxley, as a label for the formulation and acceptance of the doctrine of evolution in the natural sciences of the nineteenth century, tends to eclipse a number of important workers, of both the schools of facts and ideas, in those sciences, each of whom contributed in some measure to the growing edifice of modern biology.

One such " lesser light " was the English naturalist St. George Jackson Mivart, who died fifty years ago, on April 1st, at the age of 72. Few biologists today recall the name Mivart although in the last three decades of the nineteenth century, his was a name—and a pen—to be reckoned with as a scientific critic, writer and controversialist. Part, if not all, of this contemporary obscurity is due, I think, to the unfortunate nature of his two most important controversies. And these controversies—one leading to his excommunication from the ranks of orthodox Darwinists,

the other to his excommunication from the ranks of orthodox Catholics—were both due to his particular conception of science, and, in keeping with that conception, his attitude toward the relationship that should exist between science and religion.

Mivart had two loves—the religion to which he had been converted as a young man, and science, particularly natural science, to which he was also a convert, from the study of law. He was a passionate student of both, often permitting his passion to obscure what was for him the most important task in life, the search for Truth.

It is difficult for us today in an intellectual era of relativity, of cultural factors in perception, of the recognized imperfection of the unaided senses in the perception of reality, to consider as anything but naive a philosophy of science which emphasized the possibility of achieving real Truth through a combination of fact-collecting and rational thinking. It is difficult now to conceive of a philosophy in which the rational man can *know* by his very rationality the world about him. But this, in essence, was the basis upon which Mivart built his systems and conceived his criticisms—a relict of the halcyon days of Descartes, Newton, and the eighteenth century Age of Enlightenment.

Writing as a student of intellectual history, Randall (1940) states that:

> Perhaps the fundamental emphasis brought by Evolution into men's minds has been upon the causal analysis of specific processes of change. . . . They have rejected the ultimate goal of both Thomas and Spinoza, the contemplation of a fixed and static structure of Truth, and adopted instead the aim of investigating all the little truths which experimentation can reveal. Not that Truth which is the source of all truths, . . . but the patient, tireless, and endless search after an infinity of finite truths in our experience—this is the present-day goal of all scientific and philosophical endeavor.

With such an emphasis, Mivart had little interest. The vast and complex workings of the universe of organic life tantalized him only as it revealed an over-all unity, reducible to a fundamental cause. In actual fact, the " patient, tireless, and endless search," the accumulation of data which expanded the horizons of factual knowledge, seems to have interested him only as a means of testing the truth of his own philosophical hypotheses and demonstrating the weaknesses, if not the falsity, of those formulated by others. I do not mean to imply, however, that Mivart's detailed observational work was either biased or without value to science. On the contrary, some of his work stands as a model of

detailed description and careful workmanship. Nor can one but express admiration for the lucid and unimpassioned marshalling and presentation of facts in support of a thesis as demonstrated in his little *Man and Apes*, first published in 1873. Here was no dabbler, no arm-chair theorist or speculator; but rather a "brilliant anatomist," in the words of Leonard Huxley, and one who knew, literally, the facts of life. One could devote this entire paper—and more—to a description of Mivart's specifically scientific works, for, as a prolific writer and indefatigable worker, he produced, with varying degrees of success, no less than seventy-three scientific papers, articles, memoirs, and books, mainly on the mammals, during the 30-odd years of his publishing career from 1864 to 1896.

Mivart was not particularly interested in anthropology as such; his orientation, mirroring to some extent that of his teacher Huxley, was toward the organic world as a whole. He did, however, in an article published in the *Quarterly Review* (Mivart, 1874) make use of the recently emerging comparative cultural data of Tylor and Lubbock to strengthen his own position with respect to the uniqueness of man and his qualitative distinction from the rest of the animal kingdom. It was this article, incidentally, published in 1874, which brought to a final break the once amicable relations that had existed between Mivart and the Darwinians. Despite his general disinterestedness in man as a biological subject, Mivart did, as early as 1875, propose to write a volume on human biology to be the first of a projected series of works to comprise "a Biology which should describe the organic world in descending order from the higher to the lower forms of life."

In a letter (Mivart, 1875), dated March 9th of that year, he wrote to John Murray, the intended publisher, in answer to the latter's question as to whether Mivart's previously published *Lessons in Elementary Anatomy* would not overlap the projected work:

> The conception and treatment of the subject in Elementary Anatomy are so different from the conception I have formed and the treatment I propose for my new work that I cannot think the former will hinder but rather that it will *promote* the sale of the latter. In the Elem. Anatomy minute structure (Histology) and Physiology are entirely excluded and the greater bulk of the book is made up of facts of comparative anatomy culled from all the Vertebrate classes. In my proposed work the first volume will consist of an anatomy of Man, much more full and complete than in the little book, including physiology, histology, races of Man, his relation to time and space and to plants and other animals as his food or his enemies. Forming a zoology of Man attractive to all interested

> **in Man so considered whether or not they care for the anatomy of other animals. The second volume, also complete in itself, will be devoted to Mammals only and with that the Elem. Anatomy cannot interfere. Then as to illustrations I have received on all hands complaints of the badness of the illustrations in the last mentioned work and their uselessness on account of their small size. My desire is that the new work should be distinguished from all now known to me by the superiority of its illustrations . . .**

Although there is some evidence that this volume was completed in manuscript, for some reason it was never published and the subject of the opening volume of the series was shifted from man to the cat; and, so far as I have as yet been able to discover, the rest of the series was never completed except for an indifferent catch-all volume on various animal types published in 1893 (Mivart, 1893).

With such a background of concrete knowledge, it cannot be said that Mivart's opposition to Darwinism stemmed, as so often was—and is—the case, from an ignorance of either the doctrine itself or of the facts of biology upon which the doctrine was based. In addition to his own contributions, Mivart was an intimate of both Thomas Huxley and Alfred Wallace and, at least, a correspondent of Darwin. Through such contacts, as well as discussions with Haeckel and others of the Darwinian school, Mivart undoubtedly gained an intimate knowledge and understanding not only of the Darwinian doctrine itself but also of its implications.

In fact, it was through such an accumulation of facts and such an intimate knowledge of Darwinism that he felt constrained to attack Natural Selection as insufficient to explain all the observable phenomena. It has been maintained that Mivart opposed Darwinism because of religious bias or even through personal vanity. Such criticisms are not only unfair, but not in keeping with the facts. Originally an ardent advocate of Darwinism, it was not until 1868, nine years after the publication of the *Origin*, that Mivart began to have doubts about Natural Selection; and it was still three years more, in 1871, that his first published criticisms, *The Genesis of Species* and a review of the *Origin* in the *Quarterly Review*, appeared. There can be no doubt that his religious convictions played some part in his growing opposition to the theory he had once championed; but there can also be no doubt that the facts as he saw them were the deciding factors in his articulate opposition. In an article written late in his life (Mivart, 1897), Mivart writes movingly of this period of doubt, divided loyalty, and, finally, decision:

> It was in 1868 that difficulties as to the theory of Natural Selection began to take shape in my mind. . . . [These] were the difficulties, or rather the impossibilities, on the Darwinian system, of accounting for the origin of human intellect, and above all for its moral intuitions—not its moral *sentiments*, but its ethical *judgments*.
>
> For the rest of that year and the first half of the next I was perplexed and distressed as to what line I ought to take in a matter so important, and which more and more appeared to me one I was bound to enter upon controversially.
>
> After many painful days and much meditation and discussion my mind was made up, and I felt it my duty first of all to go straight to Professor Huxley and tell him all my thoughts, feelings, and intentions in the matter without the slightest reserve, including what it seemed to me I must do as regarded the theological aspect of the question. Never before or since have I had a more painful experience than fell to my lot in his room at the School of Mines on that 15th of June, 1869. As soon as I had made my meaning clear, his countenance became transformed as I had never seen it. Yet he looked more sad and surprised than anything else. He was kind and gentle as he said regretfully, but most firmly, that nothing so united or severed men as questions such as those I had spoken of.

Although the difficulties of natural selection to explain the origin of man, particularly as it applies to his intellect, receive increasing emphasis in his later writings, Mivart's earlier critiques of Darwinism depended primarily upon the facts of biology. In replying to a criticism of Huxley's, for instance, he stated (Mivart, 1874a) that the first object of his *Genesis of Species* "was to show that the Darwinian theory is untenable and that natural selection is not *the* origin of species. This was and is my conviction purely as a man of science, and I maintain it upon scientific grounds only."

Despite the scientific arguments mustered, Mivart's system reeked too much of the mysticism of Owen to be acceptable to the followers of Natural Selection. The unfortunate aspect of the entire dispute lies in the fact that both Darwin and Mivart were talking about and explaining different phenomena, presumably under the impression that their systems were at cross purposes. Appealing to the facts of observation, Darwin, in the *Origin*, assumed without question the existence of variation within species and the maintenance of that variation through accidental changes, neither the nature nor cause of which were adequately explained until the perfection of his hypothesis of Pangenesis. Operating within this frame of reference, he could maintain that Natural Selection, operating on such variable populations, was the cause of the origin of new species. To Mivart, however, this was not sufficient. The problem

to him was not so much one of the effect of Natural Selection once the variable population was assumed, but rather the cause of the original variations which were, *a priori*, fundamental to the production of new species. Arguing from the original uselessness of incipient organs and the independent development of similar organs, seemingly useless in any environmental situation, in different groups, he maintained that Natural Selection was not sufficient to explain the rise of variations and, consequently, merely a secondary factor in the origin of species, coming into play only after the rise of the variations themselves. For Mivart, the cause of the variations must be sought in some internal force, which then becomes the true cause of the origin of species and the guiding principle in evolution. Jealous of the rights of scientific priority, even for others, Mivart was quick to acknowledge the similarity of his system with that of Owen. But comparing his with that of the great anatomist, he writes (Mivart, 1871):

> The view propounded in this work allows, however, a greater and more important part to the share of external influences, it being believed by the author, however, that these external influences equally with the internal ones are the results of one harmonious action underlying the whole of Nature, organic and inorganic, cosmical, physical, chemical, terrestrial, vital, and social.
>
> According to this view, an internal law presides over the actions of every part of every individual, and of every organism as a unit, and of the entire organic world as a whole. It is believed that this conception of an internal innate force will ever remain necessary, however much its subordinate processes and actions may become explicable:
>
> That by such a force, from time to time, new species are manifested by ordinary generation . . . these new forms not being monstrosities but harmonious self-consistent wholes. That thus, as specific distinctness is manifested by obscure sexual conditions, so in obscure sexual modifications specific distinctions arise.
>
> That these "jumps" are considerable in comparison with the minute variations of "Natural Selection"—are in fact sensible steps, such as discriminate species from species.
>
> That the latent tendency which exists to these sudden evolutions is determined to action by the stimulus of external conditions.
>
> That "Natural Selection" removes the antecedent species rapidly when the new one evolved is more in harmony with surrounding conditions.
>
> That "Natural Selection" favours and develops useful variations, though it is impotent to originate them or to erect the physiological barrier which seems to exist between species.
>
> By some such conception as this, the difficulties here enumerated, which beset the theory of "Natural Selection" pure and simple, are to be got over.

Most of Mivart's factual arguments were never discussed, much less refuted. And in the light of the complete discrediting of the doctrine of the inheritance of acquired characters implicit in the Darwinian concept of Pangenesis and of the findings of modern genetics, who can say that Mivart was not, at least, on the right track? The tragedy of Mivart is that although he saw so clearly the imperfections of the Darwinian doctrine as a complete evolutionary system, and although he himself approached such completion in modern terms, he was unable, because of a basically theological position, to elaborate his views into what we might call today a genetic basis of evolution. One of his obituaries (*Athenaeum*, 1900) intimates as much, though unconscious of what Mivart almost achieved.

"His scientific work," it reads, "was however, limited by his religious creed. Perhaps had he been less dominated by preconceived ideas he would, with his industry, have achieved something more striking than anything that can be claimed for him."

Without his religion, however, Mivart would not be Mivart, just as he would not be without his science. The two were inextricably linked in his thinking; and it was from this linkage that Mivart derived the underlying concept of all his thinking—that of *unity*. To Mivart, there seem to have been no "mutually exclusives" in the area of truth, whether he dealt with the individual organism, the organic world, or the universe; or whether he dealt with the field of science, narrowly defined, religion, or human knowledge as a whole. Everything was united in his thinking and much of his writing deals with problems of synthesis of quite diverse orders.

He saw himself, scientist and religionist, as the unifier of the seemingly contradictory fields of science and religion. His articles and book opposing Darwinism were, in his eyes, not so much negative criticisms as positive contributions to a reconciliation of science—in this case evolution—and religion. He maintained again and again, much to the dismay of the orthodox of both camps, that the two were harmonious. He lived to be virtually rejected by the one and actually by the other.

His writings in science are replete with this concept of the unity of all life, of the necessity of taking the broad, over-all view in order to understand fully and completely the phenomena of nature. It is from the understanding of the existence of such complex interrelationships that he recognized the importance in biology of a field which he was one of the first to define. He called it Hexicology and defined it (Mivart,

1894) as "a special science . . . devoted to a study of the relations which exist between organisms and their environment as regards the nature of the localities which they frequent, the temperatures and amounts of light which suit them, and their relations to other organisms as enemies, rivals, or accidental and involuntary benefactors." Today, of course, this "special science" is Ecology. Furthermore, even in his texts, he sought a unity of presentation, rather than, as was common, a strict detailing of facts. He expresses this view best, and least formally, when, in answer to an unfriendly review of *The Cat*, he writes (Mivart, 1881) to his publisher:

"The writer has tried his very best to carp. To find fault with my treating of subjects to be found 'in plenty of acceptable books' is perverse when a man of the least intelligence must see that my object was to produce a work complete in itself with *the express intention* of saving the student from the necessity of having recourse to such other books."

A peculiarly pertinent expression of one phase of this unity was given by Mivart with reference to the nature of man. Unlike his fellow anatomists, Mivart was unwilling to define man in simple morphological terms. He writes (Mivart, 1873):

> Man being, as the mind of each man may tell him, an existence, not only conscious, but conscious of his own consciousness; one not only acting on inference, but capable of analysing the *process* of inference; a creature not only capable of acting well or ill, but of understanding the ideas "virtue" and "moral obligation," with their correlatives freedom of choice and responsibility—man being all this, it is at once obvious that the principal part of his being is his mental power. . . .
>
> We must entirely dismiss, then, the conception that mere anatomy by itself can have any decisive bearing on the question as to man's nature and being as a whole. To solve this question, recourse must be had to other studies; that is to say, to philosophy, and especially to that branch of it which occupies itself with mental phenomena—psychology.

Accepting this statement as written in 1873, even with its philosophical overtones, as a basic point of view with respect to the study of man and the solution of problems of his origin, one notes a modern touch—the necessity of defining man as a complete behavioral unit rather than a collected mass of structural details. Whatever one thinks of the origin of the mental processes which mark man off from the rest of the animal kingdom, he must recognize that mere comparative anatomy cannot itself solve the problem of man's origin and evolution. Some recourse must be had, as Mivart so well recognized, to the findings

of comparative psychology and behavior. Although Huxley recognized the necessity of taking into account "psychical phenomena" in formulating the pedigree of man, he did so fleetingly and, in fact, only as something of an after-thought. In the post-Darwinian period, among biologists, only Mivart and then Romanes seem to have recognized the necessity of dealing with the problem of behavior or of the "mind" as it was then recognized, for purposes of establishing the truth or falsity of the Darwinian concept of evolution. Both Mivart and Romanes, it should be noted, attacked the problem in a thoroughly scientific fashion although each arrived at quite different conclusions.

As religion was the fundamental base upon which Mivart built his scientific systems in his earlier years, science became the base upon which he structured his religion in his later years, although he constantly sought for a middle ground, a *tertium quid*, between the two. He was quick to strike out against the authoritarianism of either, against the acceptance through faith or fame of doctrines that could not be supported by reason.

Inevitably his attempts to reconcile the dogma of the Catholic church with his scientific knowledge led to a hampering of his own expressions in the field of reconciliation. He stated in print as early as 1885 (Mivart, 1885) that Catholics could be good Catholics and still believe in the discoveries of science—that, in fact, since in scientific matters the Church authorities had so often been in error, it was incumbent upon him who was both Catholic and scientist to elicit the truth in those areas in which the Church was apt to be fallible. Such an attitude, if widely accepted, could well be the disintegrating factor in the Catholic Church as a highly organized system of belief and cosmic explanation. For the next fifteen years, Mivart dwelt upon the same theme, of the necessity of the Church to change its beliefs in order to harmonize with the findings of science. So far did he go in this direction, that after two singularly frank articles (Mivart 1900, Mivart 1900a) published three months before his death, the then Archbishop of Westminster called upon Mivart (Vaughan, 1900) to make a full confession of faith, explicitly maintaining the truth of the Scriptural record, at the risk of excommunication. In answer to this request, Mivart wrote (Mivart, 1900b):

> All of us, however submissive to authority, must in the last resort, rest upon the judgment of our individual reason. How otherwise could we know that authority had spoken at all or what it had said? It is impossible to accept anything as true which is a contradiction in terms.

> Upon that truth all theological reasoning is based and all other reasoning also. . . . It is now evident that a vast and impassable abyss yawns between Catholic dogma and science, and no man with ordinary knowledge, can henceforth join the communion of the Roman Catholic Church if he correctly understands what its principles and its teachings really are, unless they are radically changed.

There could be no questioning his attitude; he was excommunicated. Two months later he died.

Although it is tempting to see tragedy in this last great controversy of Mivart's, it had for Mivart, himself, no such meaning. He recognized what he was doing and consciously made his peace with science, having realized that despite fifty years of effort, he could not bridge the gap which existed between nineteenth century science and nineteenth century religion. Writing a few weeks before his death to a friend, he summed up his life's work (Mivart, 1900c):

> The various articles and few books I have written have always represented my convictions at the time as accurately as I could represent them. . . . I have no more leaning to atheism or agnosticism now than I ever had; but the inscrutable, incomprehensible energy pervading the universe and (as it seems to me) disclosed by science, differs profoundly, as I read nature, from the God worshipped by Christians.

With such a view, he could finally say (Mivart, 1900b): "Happily I can now speak with entire frankness as to all my convictions. *Liberavi animam meam* [I have liberated my mind]. I can sing my *Nunc dimittis* and calmly await the future."

BIBLIOGRAPHY

Athenaeum, 1900. April 7, No. 3780. *London.*

MIVART, ST. GEORGE. 1871. On the Genesis of Species. *London.*

1873. Man and Apes. *London.*

1874. Primitive Man—Tylor and Lubbock. *Quarterly Review* (*London*), July.

1874a. Reply to T. H. Huxley. *Contemporary Review.*

1875. Letter to John Murray, in the possession of Sir John Murray, *London.*

1881. Letter to John Murray, in the possession of Sir John Murray, *London.*

1885. Modern Catholics and Scientific Freedom. *Nineteenth Century*, July, Vol. 18, pp. 30-47.

1893. Types of Animal Life. *London.*

1894. The Groundwork of Science. *London.*
See also: The Cat. *London,* 1881.

1897. Reminiscences of T. H. Huxley. *Nineteenth Century,* December, Vol. 42, pp. 985-998.

1900. The continuity of Catholicism. *Nineteenth Century,* January, Vol. 47, pp. 51-72.

1900a. Some recent apologists. *Fortnightly Review,* January.

1900b. Letter to Cardinal Vaughan printed in *The Times,* (*London*), January.

1900c. Letter to anonymous friend in *The Academy,* April 7, Vol. 58.

RANDALL, JOHN H. JR. 1940. The Making of the Modern Mind. *New York.*

VAUGHAN, CARDINAL. 1900. Letter reprinted in *The Times* (*London*), January.

JACOB W. GRUBER
Temple University